```
  AUTHORS   Sasaki,T., Miyao,A. and Yamamoto,K.
  TITLE     Rice cDNA from callus 1995
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program
            2-1-2 Kannondai,Tsukuba
            Ibaraki,
            Japan 305
            Tel: 0298-38-7441
            Fax: 0298-38-7468
            Email: tsasaki@abr.affrc.go.jp
            High quality sequence stop: 212.
FEATURES              Location/Qualifiers
     source           1. .403
                      /organism="Oryza sativa"
                      /strain="Nipponbare"
                      /db_xref="taxon:4530"
                      /clone_lib="Rice green shoot"
                      /note="Green shoot (8 days old)"
BASE COUNT       60 a    124 c    137 g     75 t      7 others
ORIGIN


  Query Match             14.4%;  Score 28;  DB 21;  Length 403;
  Best Local Similarity   53.7%;  Pred. No. 13;
  Matches   58;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy      18 GACTCAGCCCCCCGTCATAATGCCAGGTATGCAGACCGAGATCAAAACGCAGGCCACGAA 77
           | | | ||  |  | ||   | || || |||| | | ||||  |   ||| | |||  ||
Db     288 GTCGCTGCTGCTGGACAGCGTCCCCGGGATGCCGTCGGAGAAGACGTCGCCGCCCAACAA 347

Qy      78 TCTTGCGGCGAATCTTTCCGCAGTCAGAGAAAGTGCCACAGCGACGCT 125
             | ||||     | | | |||  |||||   | |  ||  ||||
Db     348 CAACTCCGCGAGGGGATTCCCCGTCGTAGAAACGTCAAAGGCCGCGCT 395


RESULT  15
AA151265
LOCUS       AA151265      425 bp    mRNA            EST       19-MAY-1997
DEFINITION  zl46e09.s1 Soares_pregnant_uterus_NbHPU Homo sapiens cDNA clone
            IMAGE:505000 3', mRNA sequence.
ACCESSION   AA151265
NID         g1719475
VERSION     AA151265.1  GI:1719475
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 425)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,
            Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
            Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
            Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J.,
            Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.
            and Marra,M.
  TITLE     Generation and analysis of 280,000 human expressed sequence tags
  JOURNAL   Genome Res. 6 (9), 807-828 (1996)
  MEDLINE   97044478
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1398063.

            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 963   Std Error: 0.00
            Seq primer: -40M13 fwd. from Amersham
            High quality sequence stop: 246.
FEATURES              Location/Qualifiers
     source           1. .425
                      /organism="Homo sapiens"
                      /db_xref="GDB:3809988"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:505000"
                      /clone_lib="Soares_pregnant_uterus_NbHPU"
                      /sex="female"
                      /dev_stage="adult"
                      /lab_host="DH10B"
                      /note="Organ: uterus; Vector: pT7T3-Pac; Site_1: Not I;
                      Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                      oligo(dT) primer [5'
                      AACTGGAAGAATTCGCGGCCGCCTTTTTTTTTTTTTTTT 3'],
                      double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Not I and cloned into the Not I
                      and Eco RI sites of the modified pT7T3 vector. Library
                      went through one round of normalization.  Library
                      constructed by M. Fatima Bonaldo."
BASE COUNT       96 a    112 c    111 g     100 t      6 others
ORIGIN


  Query Match             14.4%;  Score 28;  DB 29;  Length 425;
  Best Local Similarity   58.3%;  Pred. No. 13;
  Matches   49;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy      66 GCAGGCCACGAATCTTGCGGCGAATCTTTCCGCAGTCAGAGAAAGTGCCACAGCGACGCT 125
           |||| |||| ||||   |||  |||  |   |   || |   ||||   |  |||  ||
Db     198 GCAGTTCACGGATCTCCAGGCCCATCCTGTTGACTGCATACCCAGTGAGGCTTCGAGCCT 257

Qy     126 GTCAGGGGAAATTAAAGGCCCGCA 149
           | ||| ||  | || || || |||
Db     258 GCCAGAGGTCAATACAGCCCAGCA 281


Search completed: October  5, 1999, 17:14:01
Job time: 8684 sec
```

```
RESULT  12
AA390758/c
LOCUS       AA390758      584 bp    mRNA            EST       27-NOV-1998
DEFINITION  LD08842.5prime LD Drosophila melanogaster embryo BlueScript
            Drosophila melanogaster cDNA clone LD08842 5prime, mRNA sequence.
ACCESSION   AA390758
NID         g2790228
VERSION     AA390758.1  GI:2790228
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 584)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 17, 1998 this sequence version replaced gi:2043744.

            Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 88  row: D  column: 6
            High quality sequence stop: 365.
FEATURES             Location/Qualifiers
     source          1. .584
                     /organism="Drosophila melanogaster"
                     /db_xref="BDGP_EST:BDcln008165"
                     /db_xref="taxon:7227"
                     /clone="LD08842"
                     /clone_lib="LD Drosophila melanogaster embryo BlueScript"
                     /sex="male and female"
                     /dev_stage="0 to 24 hours mixed stage embryonic"
                     /lab_host="SOLR"
                     /note="Organ: embryo; Vector: BlueScript SK; Site_1:
                     EcoRI; Site_2: XhoI; Constructed using Stratagene ZAP-cDNA
                     Synthesis kit. Oligo dT-primed and directionally cloned at
                     EcoRI and XhoI in BlueScript SK(+/-)"
BASE COUNT      166 a    173 c    150 g     95 t
ORIGIN
```

```
  Query Match           14.5%;  Score 28.2;  DB 38;  Length 584;
  Best Local Similarity  49.0%;  Pred. No. 13;
  Matches   75;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;

Qy      23 AGCCCCCCGTCATAATGCCAGGTATGCAGACCGAGATCAAAACGCAGGCCACGAATCTTG 82
           || |  || ||    |||||         | ||  |||   || | |  | |  |    |
Db     315 AGGCTACCCTCGGGCTGCCATCAGGCGTGTCCAGGATTTCAATGGACTCGATTATGGATT 256

Qy      83 CGGCGAATCTTTCCGCAGTCAGAGAAAGTGCCACAGCGACGCTGTCAGGGGAAATTAAAG 142
           ||||    |  | || | ||  | |    | || ||  |||| || |  | ||| | ||
Db     255 GGGCGGGATTCGCGGCGGGCAAGGGACCAACGACTGCCTCGCTTTCTGCAGCAATAACAG 196

Qy     143 GCCCGCAACTGGAAGATTTTCCCGCGCTGATCA 175
           || | || || |    |||    |  ||||
Db     195 ATCCCCCACCGGTACCGCTTCCATGTCCCATCA 163
```

```
RESULT  13
AA828848
LOCUS       AA828848      306 bp    mRNA            EST       07-APR-1998
DEFINITION  od80d11.s1 NCI_CGAP_Ov2 Homo sapiens cDNA clone IMAGE:1374261
            similar to TR:Q15417 Q15417 ACIDIC CALPONIN. ;, mRNA sequence.
ACCESSION   AA828848
NID         g2901947
VERSION     AA828848.1  GI:2901947
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 306)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2045799.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: David B. Krizman, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 585    Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham.
FEATURES             Location/Qualifiers
     source          1. .306
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="2; 3q"
                     /clone="IMAGE:1374261"
                     /clone_lib="NCI_CGAP_Ov2"
                     /sex="female"
                     /tissue_type="ovary"
                     /lab_host="DH10B"
                     /note="Vector: pAMP10; mRNA made from invasive ovarian
                     tumor, cDNA made by oligo-dT priming.  Non-directionally
                     cloned. Size-selected on agarose gel, average insert  size
                     600 bp. Reference:  Krizman et al. (1996) Cancer Research
                     56:5380-5383."
BASE COUNT       95 a     67 c     63 g     81 t
ORIGIN
```

```
  Query Match           14.5%;  Score 28.2;  DB 39;  Length 306;
  Best Local Similarity  64.6%;  Pred. No. 9.5;
  Matches   42;  Conservative    0;  Mismatches   23;  Indels    0;  Gaps    0;

Qy      48 GCAGACCGAGATCAAAACGCAGGCCACGAATCTTGCGGCGAATCTTTCCGCAGTCAGAGA 107
           || |  |||  ||| |||| | |||| | | |  | || |||  ||| | | |||| ||
Db      35 GCCGGTCGACTTCACAACGGAAGCCAAGGAACAGGAAGCAAATGGTTCGGAAATCAGTGA 94

Qy     108 AAGTG 112
           ||||
Db      95 TAGTG 99
```

```
RESULT  14
D48418
LOCUS       D48418        403 bp    mRNA            EST       02-AUG-1995
DEFINITION  RICS14604A Rice green shoot Oryza sativa cDNA, mRNA sequence.
ACCESSION   D48418
NID         g702127
VERSION     D48418.1  GI:702127
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1  (bases 1 to 403)
```

```
                     /db_xref="taxon:10090"
                     /map="11q23"
                     /clone="1500035I19"
                     /clone_lib="Mus musculus adult C57BL/6J cerebellum"
                     /sex="male"
                     /tissue_type="cerebellum"
                     /dev_stage="adult"
BASE COUNT       71 a     47 c     64 g    110 t
ORIGIN
```

Query Match 14.9%; Score 29; DB 50; Length 292;
Best Local Similarity 53.0%; Pred. No. 4.9;
Matches 62; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

```
Qy      11 GTGTAAGGACTCAGCCCCCCGTCATAATGCCAGGTATGCAGACCGAGATCAAAACGCAGG 70
           | || ||    |||       | ||||| || | ||||   |  |||   | ||| || |
Db     292 GGGTCAGTTAACAGTTAATTTTAATAATACCTGTTATGTTTAAAGAGTCAACAACACACG 233

Qy      71 CCACGAATCTTGCGGCGAATCTTTCCGCAGTCAGAGAAAGTGCCACAGCGACGCTGT 127
             ||   | | | |||| ||| | ||||  ||  |  |||   | | | || ||
Db     232 AAGAAAAGTCTTCTGTGAATGTTTACACAGTGTGAACACATGCTTTAACTATGCAGT 176
```

```
RESULT  10
AA297724/c
LOCUS       AA297724      404 bp    mRNA            EST       18-APR-1997
DEFINITION  EST113280 Infant brain Homo sapiens cDNA 3' end, mRNA sequence.
ACCESSION   AA297724
NID         g1950058
VERSION     AA297724.1  GI:1950058
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 404)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fields,C. and Venter,J.C.
  TITLE     3,400 expressed sequence tags identify diversity of transcripts
            from human brain
  JOURNAL   Nature Genet. 4, 256-267 (1993)
  MEDLINE   93364420
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1326794.
            Other_ESTs: EST56078 THC170768
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13-21.
FEATURES             Location/Qualifiers
     source          1. .404
                     /organism="Homo sapiens"
                     /db_xref="ATCC (inhost):150490"
                     /db_xref="taxon:9606"
                     /map="7 pter-cen; 17; 17q11.2-q23"
                     /clone_lib="Infant brain"
                     /sex="female"
                     /dev_stage="infant"
                     /note="Organ: brain; Vector: lafmid BA; Site_1: HindIII;
                     Site_2: NotI"
BASE COUNT       88 a    109 c    120 g     85 t      2 others
ORIGIN
```

Query Match 14.7%; Score 28.6; DB 31; Length 404;
Best Local Similarity 55.6%; Pred. No. 7.8;
Matches 55; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

```
Qy      13 GTAAGGACTCAGCCCCCCGTCATAATGCCAGGTATGCAGACCGAGATCAAAACGCAGGCC 72
           || | |  |||| | || |  |  ||  ||| | ||| ||| |||  |||   ||  ||
Db     248 GTGATGGTTCAGGCTCCTGGGAAGATTGCAGCTTTGCTGACGGAGGCCAACCTGCCTGCT 189

Qy      73 ACGAATCTTGCGGCGAATCTTTCCGCAGTCAGAGAAAGT 111
             |   || |||  |   |  | || | |||  ||| | |
Db     188 CCTGCTCCTGCTTCTGCTTCTGCCCCTGTCCTAGAGACT 150
```

```
RESULT  11
AI728053
LOCUS       AI728053      620 bp    mRNA            EST       11-JUN-1999
DEFINITION  BNLGHi9696 Six-day Cotton fiber Gossypium hirsutum cDNA 5' similar
            to (AF051898) coronin binding protein [Dictyostelium discoideum],
            mRNA sequence.
ACCESSION   AI728053
NID         g5046839
VERSION     AI728053.1  GI:5046839
KEYWORDS    EST.
SOURCE      upland cotton.
  ORGANISM  Gossypium hirsutum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
REFERENCE   1  (bases 1 to 620)
  AUTHORS   Blewitt,M., Matz,E.C., Davy,D.F. and Burr,B.
  TITLE     ESTs from developing cotton fiber
  JOURNAL   Unpublished (1999)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3187769.

            Contact: Ben Burr
            Biology Department
            Brookhaven National Laboratory
            Upton, NY 11973, USA
            Tel: 516-344-3396
            Fax: 516-344-3407
            Email: burr@bnlux1.bnl.gov
            Seq primer: T3 Primer.
FEATURES             Location/Qualifiers
     source          1. .620
                     /organism="Gossypium hirsutum"
                     /cultivar="Acala Maxxa"
                     /db_xref="taxon:3635"
                     /map="17q21; 8q13.2"
                     /clone_lib="Six-day Cotton fiber"
                     /tissue_type="immature fiber"
                     /dev_stage="Six days post anthesis"
                     /lab_host="XL1-Blue"
                     /note="Vector: pBluescript II KS+"
BASE COUNT      259 a    157 c    122 g     82 t
ORIGIN
```

Query Match 14.7%; Score 28.6; DB 51; Length 620;
Best Local Similarity 47.1%; Pred. No. 9.5;
Matches 88; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

```
Qy       3 GGTTACAAGTGTAAGGACTCAGCCCCCCGTCATAATGCCAGGTATGCAGACCGAGATCAA 62
           ||| | ||    ||  |  |  | ||  || | |  | ||   |||   |   || |||
Db      71 GGTCAGAACGAGAACAATGCCACTCCAAGTAACACCGACAACAATGAGAATGCAGGTCAG 130

Qy      63 AACGCAGGCCACGAATCTTGCGGCGAATCTTTCCGCAGTCAGAGAAAGTGCCACAGCGAC 122
           ||||    | | |   ||    || |  ||  |  |||  | || ||   | ||| |||
Db     131 AACGAGAACAATGTGACTCAGAGCAACACTGACAACAGCAACAGCAACAACAACAACGAA 190

Qy     123 GCTGTCAGGGGAAATTAAAGGCCCGCAACTGGAAGATTTTCCCGCGCTGATCAAACAGGC 182
             ||  ||   |||  | |     ||  ||   ||   |  |  |  |||  |  || |
Db     191 AATGCAAGTCAAAACGAGAACAATGCTGCTCAGAGCAGTAACAACAATGAAAATGCAAGT 250

Qy     183 GAGTCTG 189
             ||  ||
Db     251 CAGGATG 257
```

```
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:635802
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 471.
FEATURES                Location/Qualifiers
     source             1. .476
                        /organism="Mus musculus"
                        /strain="C57BL/6J"
                        /db_xref="taxon:10090"
                        /map="19; 21"
                        /clone="IMAGE:1177954"
                        /clone_lib="Soares mouse mammary gland NbMMG"
                        /sex="male"
                        /tissue_type="mammary gland"
                        /dev_stage="4 weeks"
                        /lab_host="DH10B"
                        /note="Organ: mammary gland; Vector: pT7T3D-Pac
                        (Pharmacia) with a modified polylinker; Site_1: Not I;
                        Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                        oligo(dT) primer [5'
                        TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATGGTTTTTTTTTTTTTTTTTTTTT
                        T 3']; double-stranded cDNA was ligated to Eco RI
                        adaptors (Pharmacia), digested with Not I and cloned into
                        the Not I and Eco RI sites of the modified pT7T3 vector.
                        RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
                        constructed and normalized by Bento Soares and M.Fatima
                        Bonaldo."
BASE COUNT      117 a     95 c    112 g    152 t
ORIGIN


  Query Match             14.9%;  Score 29;  DB 37;  Length 476;
  Best Local Similarity   55.4%;  Pred. No. 6.1;
  Matches   56;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy      32 TCATAATGCCAGGTATGCAGACCGAGATCAAAACGCAGGCCACGAATCTTGCGGCGAATC 91
           | ||||| || | ||||   |  |||   | ||| || |     ||   | | | ||||
Db     457 TAATAATACCTGTTATGTTTAAAGAGTCAACAACACACGAGGAAAAGTCTTCTGTGAATG 398

Qy      92 TTTCCGCAGTCAGAGAAAGTGCCACAGCGACGCTGTCAGGG 132
           ||| | ||||  ||  |  |||   | | | || ||   ||
Db     397 TTTACACAGTGTGAACACATGCTTTAACTATGCAGTGGAGG 357


RESULT   8
AU005963
LOCUS       AU005963      699 bp    mRNA            EST       19-JAN-1999
DEFINITION  AU005963 Bombyx mori p50(Daizo) Bombyx mori cDNA clone wv40308,
            mRNA sequence.
ACCESSION   AU005963
NID         g4163347
VERSION     AU005963.1  GI:4163347
KEYWORDS    EST.
SOURCE      domestic silkworm.
  ORGANISM  Bombyx mori
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;
            Ditrysia; Bombycoidea; Bombycidae; Bombyx.
REFERENCE   1  (bases 1 to 699)
  AUTHORS   Mita,K., Morimyo,M., Shimada,T., Okano,K. and Maeda,S.
  TITLE     Establishment of cDNA database of Bombyx mori
  JOURNAL   Unpublished (1999)
COMMENT     On Jan 17, 1998 this sequence version replaced gi:1900599.

            Contact: Mita K
            Genome Research Group
            National Institute of Radiological Sciences
            Anagawa 4-9-1, Inage, Chiba 263-8555, Japan
            Email: kmita@uexs64.nirs.go.jp
            PROJECT = 'CREST project by JST'.
FEATURES                Location/Qualifiers
     source             1. .699
                        /organism="Bombyx mori"
                        /strain="p50(Daizo)"
                        /db_xref="taxon:7091"
                        /map="19; 21"
                        /clone="wv40308"
                        /clone_lib="Bombyx mori p50(Daizo)"
BASE COUNT      289 a    162 c    136 g    112 t
ORIGIN


  Query Match             14.9%;  Score 29;  DB 45;  Length 699;
  Best Local Similarity   61.0%;  Pred. No. 7.3;
  Matches   47;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy      28 CCCGTCATAATGCCAGGTATGCAGACCGAGATCAAAACGCAGGCCACGAATCTTGCGGCG 87
           || |  ||| || | | |||  |||| | | ||| ||| || | || |||    | | |
Db     262 CCGGAAATAGTGTCCGTTATCGAGACAGTGGTCATAACCCAAGTCATGAACAAAGAGACC 321

Qy      88 AATCTTTCCGCAGTCAG 104
           |   |||||| ||  ||
Db     322 AGCGTTTCCGAAGAGAG 338


RESULT   9
AV032477/c
LOCUS       AV032477      292 bp    mRNA            EST       03-JUN-1999
DEFINITION  AV032477 Mus musculus adult C57BL/6J cerebellum Mus musculus cDNA
            clone 1500035I19, mRNA sequence.
ACCESSION   AV032477
NID         g4832025
VERSION     AV032477.1  GI:4832025
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 292)
  AUTHORS   Carninci,P., Shibata,K., Ozawa,Y., Konno,H., Ito,M., Aizawa,K.,
            Akahira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T.,
            Hara,A., Hayatsu,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M.,
            Kawai,J., Kikuchi,N., Kojima,Y., Matsuyama,T., Niitsuma,H., Oda,H.,
            Owa,C., Sato,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y.,
            Sugahara,Y., Suzuki,H., Suzuki,H., Tateno,M., Tomaru,Y.,
            Tominaga,N., Watanabe,S., Yagame,M., Yamamura,T., Yokota,T.,
            Yoshino,M., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs
  JOURNAL   Unpublished (1999)
COMMENT     On Jun 22, 1998 this sequence version replaced gi:3247064.

            Contact: Chie Owa
            Genome Science Laboratory
            RIKEN
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-9145
            Fax: 81-298-36-9098
            Email: genome-res@rtc.riken.go.jp
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length cDNA
            (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
            Transcriptional sequencing: A method for DNA sequencing using RNA
            polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
            Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES                Location/Qualifiers
     source             1. .292
                        /organism="Mus musculus"
                        /strain="C57BL/6J"
```

```
VERSION     AI649538.1  GI:4730372
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Zea.
REFERENCE   1  (bases 1 to 340)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     On Mar 10, 1998 this sequence version replaced gi:2948518.

            Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 603006  row: A  column: 03.
FEATURES             Location/Qualifiers
     source          1. .340
                     /organism="Zea mays"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /clone_lib="603 - stressed root cDNA library from
                     Wang/Bohnert lab"
                     /tissue_type="seedling"
                     /dev_stage="salt stress"
                     /lab_host="E. coli XL Gold"
                     /note="Organ: root; Vector: pBluescriptII SK(+) XR;
                     Seedling stressed root cDNA library from Wang/Bohnert lab"
BASE COUNT       88 a     74 c     52 g    125 t      1 others
ORIGIN
```

```
  Query Match             15.1%;  Score 29.2;  DB 49;  Length 340;
  Best Local Similarity   57.8%;  Pred. No. 4.5;
  Matches   52;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy      99 AGTCAGAGAAAGTGCCACAGCGACGCTGTCAGGGGAAATTAAAGGCCCGCAACTGGAAGA 158
           ||  ||||| || |  | || ||  | | | | ||||  ||     || || |   |||
Db     302 AGAGAGAGAGAGAGAGAGAGAGTCCGACTGAGGAACATACCACACCACACCACCAAGC 243

Qy     159 TTTTCCCGCGCTGATCAAACAGGCGAGTCT 188
           || |||  || || || | || || |  ||
Db     242 TTCTCCACCGTTGCTCGACCATGCTATGCT 213
```

```
RESULT   6
AV008387/c
LOCUS       AV008387      260 bp    mRNA            EST       03-JUN-1999
DEFINITION  AV008387 Mus musculus 18-day embryo C57BL/6J Mus musculus cDNA
            clone 1110011C06, mRNA sequence.
ACCESSION   AV008387
NID         g4785374
VERSION     AV008387.1  GI:4785374
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 260)
  AUTHORS   Carninci,P., Shibata,K., Ozawa,Y., Konno,H., Ito,M., Aizawa,K.,
            Akahira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T.,
            Hara,A., Hayatsu,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M.,
            Kawai,J., Kikuchi,N., Kojima,Y., Matsuyama,T., Niitsuma,H., Oda,H.,
            Owa,C., Sato,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y.,
            Sugahara,Y., Suzuki,H., Suzuki,H., Tateno,M., Tomaru,Y.,
            Tominaga,N., Watanabe,S., Yagame,M., Yamamura,T., Yokota,T.,
            Yoshino,M., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs
  JOURNAL   Unpublished (1999)
COMMENT     On May 18, 1998 this sequence version replaced gi:3138038.

            Contact: Chie Owa
            Genome Science Laboratory
            RIKEN
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-9145
            Fax: 81-298-36-9098
            Email: genome-res@rtc.riken.go.jp
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length cDNA
            (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
            Transcriptional sequencing: A method for DNA sequencing using RNA
            polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
            Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .260
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /map="9 q21-q33"
                     /clone="1110011C06"
                     /clone_lib="Mus musculus 18-day embryo C57BL/6J"
                     /sex="mixed"
                     /dev_stage="18-day embryo"
BASE COUNT       72 a     47 c     43 g     98 t
ORIGIN
```

```
  Query Match             15.1%;  Score 29.2;  DB 49;  Length 260;
  Best Local Similarity   52.5%;  Pred. No. 4;
  Matches   64;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy      11 GTGTAAGGACTCAGCCCCCCGTCATAATGCCAGGTATGCAGACCGAGATCAAAACGCAGG 70
           | || ||    |||       | ||||| || | ||||   |  |||   | ||| || |
Db     256 GGGTCAGTTAACAGTTAATTTTAATAATACCTGTTATGTTTAAAGAGTCAACAACACACG 197

Qy      71 CCACGAATCTTGCGGCGAATCTTTCCGCAGTCAGAGAAAGTGCCACAGCGACGCTGTCAG 130
             ||   | | | |||| ||| | ||||  ||  |  |||   | | | || ||
Db     196 AGGAAAAGTCTTCTGTGAATGTTTACACAGTGTGAACACATGCTTTAACTATGCAGTGGA 137

Qy     131 GG 132
           ||
Db     136 GG 135
```

```
RESULT   7
AA718551/c
LOCUS       AA718551      476 bp    mRNA            EST       29-DEC-1997
DEFINITION  vt85e06.r1 Soares mouse mammary gland NbMMG Mus musculus cDNA clone
            IMAGE:1177954 5', mRNA sequence.
ACCESSION   AA718551
NID         g2730825
VERSION     AA718551.1  GI:2730825
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 476)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Nov 6, 1997 this sequence version replaced gi:931960.

            Contact: Marra M/Mouse EST Project
```

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 424)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 5, 1998 this sequence version replaced gi:2581140.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
            Ph.D., Gerald Marti, M.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 395.
FEATURES            Location/Qualifiers
     source         1. .424
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /map="1p; 21q"
                    /clone="IMAGE:1356675"
                    /clone_lib="NCI_CGAP_GCB1"
                    /tissue_type="germinal center B cell"
                    /lab_host="DH10B"
                    /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                    polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                    was prepared from human tonsillar cells enriched for
                    germinal center B cells by flow sorting (CD20+, IgD-),
                    provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                    (NCI) and Dr. Gerald Marti (CBER).  cDNA synthesis was
                    primed with a Not I - oligo(dT) primer
                    [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTT-
                    3']. Double-stranded cDNA was ligated to Eco RI adaptors
                    (Pharmacia), digested with Not I and cloned into the Not I
                    and Eco RI sites of the modified pT7T3 vector. Library
                    went through one round of normalization, and was
                    constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      129 a     65 c     92 g    138 t
ORIGIN


  Query Match             15.2%;  Score 29.4;  DB 38;  Length 424;
  Best Local Similarity   66.7%;  Pred. No. 4.2;
  Matches   42;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy      47 TGCAGACCGAGATCAAAACGCAGGCCACGAATCTTGCGGCGAATCTTTCCGCAGTCAGAG 106
           ||||||| |  | |||| ||    || ||| |||| ||   ||  |||  | |||| ||
Db     197 TGCAGACCTACTTTAAAAAGCTTTTCAAGAAGCTTGAGGTTGATAGTTCTTCGGTCATAG 256

Qy     107 AAA 109
           |||
Db     257 AAA 259


RESULT   4
AA846374/c
LOCUS       AA846374       326 bp    mRNA            EST       04-MAR-1998
DEFINITION  ai84h07.s1 Soares_parathyroid_tumor_NbHPA Homo sapiens cDNA clone
            IMAGE:1387549 3', mRNA sequence.
ACCESSION   AA846374
NID         g2932514
VERSION     AA846374.1  GI:2932514
KEYWORDS    EST.
```

```
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 326)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2150835.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 310.
FEATURES            Location/Qualifiers
     source         1. .326
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /map="21q"
                    /clone="IMAGE:1387549"
                    /clone_lib="Soares_parathyroid_tumor_NbHPA"
                    /tissue_type="parathyroid tumor"
                    /dev_stage="adult"
                    /lab_host="DH10B (ampicillin resistant)"
                    /note="Organ: parathyroid gland; Vector: pT7T3D
                    (Pharmacia) with a modified polylinker; Site_1: Not I;
                    Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                    oligo(dT) primer
                    [5'-
                    TGTTACCAATCTGAAGTGGGAGCGGCCGCACCAATTTTTTTTTTTTTTTTTTTTTT
                    T-3'], double-stranded cDNA was size selected, ligated to
                    Eco RI adapters (Pharmacia), digested with Not I and
                    cloned into the Not I and Eco RI sites of a modified pT7T3
                    vector (Pharmacia). Library went through one round of
                    normalization to a Cot = 5. Library constructed by Bento
                    Soares and M.Fatima Bonaldo. RNA from sporadic parathyroid
                    adenomas was kindly provided by Dr. Stephen Marx, National
                    Institute of Diabetes and Digestive and Kidney Diseases,
                    NIH."
BASE COUNT      75 a     82 c    101 g     67 t      1 others
ORIGIN


  Query Match             15.1%;  Score 29.2;  DB 39;  Length 326;
  Best Local Similarity   55.6%;  Pred. No. 4.4;
  Matches   55;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy      13 GTAAGGACTCAGCCCCCCGTCATAATGCCAGGTATGCAGACCGAGATCAAAACGCAGGCC 72
           || | |  |||| | || |  |  ||  ||| | ||| ||| |||  |||   ||  ||
Db     254 GTGATGGTTCAGGCTCCTGGGAAGATNGCAGCTTTGCTGACGGAGGCCAACCTGCCTGCT 195

Qy      73 ACGAATCTTGCGGCGAATCTTTCCGCAGTCAGAGAAAGT 111
             |   || |||  |   |  | || | |||  ||| | |
Db     194 CCTGCTCCTGCTTCTGCTTCTGCCCCTGTCCTAGAGACT 156


RESULT   5
AI649538/c
LOCUS       AI649538       340 bp    mRNA            EST       30-APR-1999
DEFINITION  603006A03.x1 603 - stressed root cDNA library from Wang/Bohnert lab
            Zea mays cDNA, mRNA sequence.
ACCESSION   AI649538
NID         g4730372
```

KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 513)
AUTHORS Newman,T., deBruijn,F.J., Green,P., Keegstra,K., Kende,H., McIntosh,L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomashow,M., Retzel,E. and Somerville,C.
TITLE Genes galore: a summary of methods for accessing results from large-scale partial sequencing of anonymous Arabidopsis cDNA clones
JOURNAL Plant Physiol. 106, 1241-1255 (1994)
MEDLINE 95148729
COMMENT On Apr 14, 1993 this sequence version replaced gi:837767.

Contact: Thomas Newman
MSU-DOE Plant Research Laboratory
Michigan State University
MSU-DOE-PRL, Michigan State University,Plant Biology Bldg.,E. Lansing,Mi
Tel: 517-353-0854
Fax: 517-353-9168
Email: 22313tcn@ibm.cl.msu.edu
Seq primer: T7 dye primer.

FEATURES Location/Qualifiers
source 1. .513
/organism="Arabidopsis thaliana"
/strain="var columbia"
/db_xref="taxon:3702"
/clone="172A1T7"
/clone_lib="Lambda-PRL2"
/note="Vector: lambda Zip-Lox; Site_1: Sal; Site_2: Not; Lambda PRL2 is a cDNA library derived from equal quantities of 4 pools of mRNA. The mRNA sources were 1) 7 day germinated etiolated seedlings; 2) tissue culture grown roots; 3) staged plants half with 24 hour light cycle, half on 16 hr light, 8 hour dark- rosettes; 4) same plants as 3 but aerial tissue (stems, flowers and siliques. The vector is BRL's lambda Zip-Lox. The cDNA inserts were directionally cloned with Sal-Not arms using oligo dT primed cDNA. "

BASE COUNT 122 a 115 c 107 g 149 t 20 others
ORIGIN

Query Match 15.3%; Score 29.6; DB 23; Length 513;
Best Local Similarity 59.5%; Pred. No. 4;
Matches 50; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

```
Qy      74 CGAATCTTGCGGCGAATCTTTCCGCAGTCAGAGAAAGTGCCACAGCGACGCTGTCAGGGG 133
           ||| | || || ||||| | |||| | ||  ||||  ||  | |   || | |||   |
Db     255 CGATTGTTCCGTCGAATGGTCCCGCCGGCATTGAAACGGCGGCGGATCCGATTTCAAATG 196

Qy     134 AAATTAAAGGCCCGCAACTGGAAG 157
           ||| |||||     ||   ||||
Db     195 GAATGAAAGGGATTGAAGATGAAG 172
```

RESULT 2
AA509689/c
LOCUS AA509689 339 bp mRNA EST 08-JUL-1997
DEFINITION vg20b10.r1 Soares mouse NbMH Mus musculus cDNA clone IMAGE:861883 5', mRNA sequence.
ACCESSION AA509689
NID g2247543
VERSION AA509689.1 GI:2247543
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 339)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1392896.

Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:505971
Seq primer: -28m13 rev2 ET from Amersham.

FEATURES Location/Qualifiers
source 1. .339
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:861883"
/clone_lib="Soares mouse NbMH"
/sex="male"
/tissue_type="heart"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCGAAAGTTTTTTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. RNA provided by Dr. Minoru Ko, Wayne State Univ. Library constructed and normalized by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 44 a 133 c 71 g 91 t
ORIGIN

Query Match 15.3%; Score 29.6; DB 34; Length 339;
Best Local Similarity 49.4%; Pred. No. 3.3;
Matches 77; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

```
Qy      34 ATAATGCCAGGTATGCAGACCGAGATCAAAACGCAGGCCACGAATCTTGCGGCGAATCTT 93
           | |  | ||| ||  | ||   ||    | || ||||  | ||||     || |||||
Db     258 AAACAGACAGATAGACGGATAAAGGGACAGACACAGGAGAAGAATGACAGGGGGAATCAG 199

Qy      94 TCCGCAGTCAGAGAAAGTGCCACAGCGACGCTGTCAGGGGAAATTAAAGGCCCGCAACTG 153
             |  |     | | | |    ||| |||   |   | ||       ||||| || | |
Db     198 AGTGGTGGAGAGGGAGGCGAGGTAGCAACGAGTTTGTGAGACTCAGCTGGCCCACAGCAG 139

Qy     154 GAAGATTTTCCCGCGCTGATCAAACAGGCGAGTCTG 189
           | | |  |||  |||| |||    |  ||| |||
Db     138 CAGGCTAACCCCAGGCTGCTCAGCAGGCAGAGCCTG 103
```

RESULT 3
AA741177
LOCUS AA741177 424 bp mRNA EST 15-JAN-1998
DEFINITION oc87d02.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1356675 3', mRNA sequence.
ACCESSION AA741177
NID g2779769
VERSION AA741177.1 GI:2779769
KEYWORDS EST.
SOURCE human.

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 1999, 17:13:57 ; Search time 1495.71 Seconds
(without alignments)
255.846 Million cell updates/sec

Title: US-09-068-804-4
Perfect score: 194
Sequence: 1 ATGGTTACAAGTGTAAGGAC.....AAACAGGCGAGTCTGGATGC 194

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 986266752 residues

Database : EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	29.6	15.3	513	23	R65129	R65129 13633 Lambd
c	2	29.6	15.3	339	34	AA509689	AA509689 vg20b10.r
	3	29.4	15.2	424	38	AA741177	AA741177 oc87d02.s
c	4	29.2	15.1	326	39	AA846374	AA846374 ai84h07.s
c	5	29.2	15.1	340	49	AI649538	AI649538 603006A03
c	6	29.2	15.1	260	49	AV008387	AV008387 AV008387
c	7	29	14.9	476	37	AA718551	AA718551 vt85e06.r
	8	29	14.9	699	45	AU005963	AU005963 AU005963
c	9	29	14.9	292	50	AV032477	AV032477 AV032477
c	10	28.6	14.7	404	31	AA297724	AA297724 EST113280
	11	28.6	14.7	620	51	AI728053	AI728053 BNLGHi969
c	12	28.2	14.5	584	38	AA390758	AA390758 LD08842.5
	13	28.2	14.5	306	39	AA828848	AA828848 od80d11.s
	14	28	14.4	403	21	D48418	D48418 RICS14604A
	15	28	14.4	425	29	AA151265	AA151265 zl46e09.s
	16	27.8	14.3	322	21	F09977	F09977 HSC39A042 n
	17	27.8	14.3	471	23	H16950	H16950 ym34a11.s1
c	18	27.8	14.3	602	27	AA033789	AA033789 zi05e05.r
	19	27.8	14.3	249	31	AA284499	AA284499 zt20b11.s
	20	27.8	14.3	432	33	AA417258	AA417258 zu07f02.s
	21	27.8	14.3	220	33	AA419343	AA419343 zv40h11.s
	22	27.8	14.3	265	33	AA428863	AA428863 zw57h01.s
	23	27.8	14.3	468	34	AA460263	AA460263 zx67e01.s
	24	27.8	14.3	695	49	AI648345	AI648345 AEMTAE36
	25	27.6	14.2	358	28	AA060191	AA060191 mj65e07.r
	26	27.6	14.2	425	30	AA232497	AA232497 zr28f07.r
c	27	27.6	14.2	822	38	AA391400	AA391400 LD10287.5
c	28	27.6	14.2	811	44	AI267130	AI267130 uj11c06.x
c	29	27.6	14.2	529	48	AI562661	AI562661 TENS2645
	30	27.6	14.2	545	48	AI595472	AI595472 mj65e07.y
	31	27.4	14.1	402	27	AA001209	AA001209 zh81c04.r
c	32	27.4	14.1	615	36	AA605983	AA605983 fa26a01.s
c	33	27.4	14.1	300	41	C95076	C95076 C95076 Pig
	34	27.4	14.1	454	44	AI323042	AI323042 mj38f03.y
	35	27.2	14.0	473	28	AA078957	AA078957 zm92c07.s
c	36	27.2	14.0	530	35	AA543288	AA543288 vj84e09.s
c	37	27.2	14.0	587	44	AU045018	AU045018 AU045018
	38	27	13.9	465	22	R26271	R26271 yh41d02.s1
	39	27	13.9	337	30	AA218794	AA218794 zq97c03.r
c	40	27	13.9	272	38	AA808455	AA808455 ob49e07.s
c	41	27	13.9	334	51	AI715989	AI715989 UI-R-Y0-a
	42	26.8	13.8	350	32	AA376458	AA376458 EST88879
	43	26.8	13.8	413	34	AA458830	AA458830 aa26a08.s
	44	26.8	13.8	315	48	AI556364	AI556364 UI-R-C2p-
c	45	26.6	13.7	457	34	AA475794	AA475794 vh24a11.r

ALIGNMENTS

RESULT 1
R65129/c
LOCUS R65129 513 bp mRNA EST 09-JAN-1998
DEFINITION 13633 Lambda-PRL2 Arabidopsis thaliana cDNA clone 172A1T7, mRNA sequence.
ACCESSION R65129
NID g937616
VERSION R65129.1 GI:937616

```
2693 CTTCGCGAATCCGAGGCGCGCGCTGCGGAGCTGGCGAGCCAGCTG..... 2737

 329 tLysAsnSerValThrValGlyGlyIleAlaGlyAlaSerGlyGlnTyrA 346
                              |||   |||:::::::::   :
2738 .......................GAGGCCACTGCTGCTGCGAAGATGT 2762

 346 laAlaThrGlnGluArgSer.................GluGlnGlnIle 356
     ::|||   |||:::|||                    |||||||||:::
2763 CAGCGGAGCAGGACCGCGAGAACACGAGGGCCACGCTAGAGCAGCAGCTT 2812

 357 SerGlnValAsnAsnArgValAlaSerThrAlaSer.............. 368
       :::   ::::::|||:::|||      ||||||
2813 CGTGACTCCGAGGAGCGCGCTGCGGAGCTGGCGAGCCAGCTGGAGTCCAC 2862

 369 .....................AspGluAlaArgGluSerSerArgLysS 378
                          :::::::::||||||||||:::|||
2863 TACTGCTGCGAAGATGTCAGCGGAGCAGGACCGCGAGAGCACGAGG.... 2908

 378 erThrSerLeuIleGlnGluMetLeuLysThrMetGluSerIleAsnGln 394
       :::::::|||   |||::::::      :::   |||       ::::::
2909 ..GCCACGCTAGAGCAGCAGCTTCGTGACTCCGAGGAGCGCGCTGCGGAG 2956

 395 SerLysAlaSerAlaLeuAlaAlaIleAlaGlyAsnIleArgAla*** 410
         :::         :::::::::|||::::::::::   ||||||
2957 CTGGCGAGCCAGCTGGAGTCCACTACTGCTGCGAAGATGTCAGCGGAG 3004
```

This Page Blank (uspto)

```
     |||                    |||:::::::::::::::::::::::
  1063 TAT.....................GAAGATCATCTTAATGAGACAAAAAC 1091

   361 nArgValAlaSerThrAlaSerAspGluAlaArgGluSerSerArgLysS 378
       ::::
  1092 TAAA........................................... 1095

   378 erThrSerLeuIleGlnGluMetLeuLysThrMetGluSerIleAsnGln 394
               ::::::||||||   :::   |||:::::::::::::
  1096 ..........CTGGATGAAATGACTCGCCTCATGAATGACCTCACTACT 1134

   395 SerLysAla 397
         |||:::
  1135 CAAAAGACA 1143

seq_name: gb_pat:I11729

seq_documentation_block:
LOCUS       I11729       3319 bp    DNA             PAT       03-MAY-1995
DEFINITION  Sequence 2 from Patent US 5411865.
ACCESSION   I11729
NID         g909481
VERSION     I11729.1  GI:909481
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 3319)
  AUTHORS   Reed,S.
  TITLE     Method of detecting anti-leishmania parasite antibodies
  JOURNAL   Patent: US 5411865-A 2 02-MAY-1995;
FEATURES             Location/Qualifiers
     source          1. .3319
                     /organism="unknown"
BASE COUNT      669 a    969 c   1196 g    485 t
ORIGIN


alignment_scores:
           Quality:  109.50            Length:    516
             Ratio:  0.429               Gaps:     22
 Percent Similarity:  49.419   Percent Identity: 19.574

alignment_block:
 US-09-068-804-6 x I11729   ..

 Align seg 1/1  to: I11729  from: 1  to: 3319

       2 LeuIleSerAsnValGlyIleAsnProAlaAlaTyrLeuAsnAsnHisSe 18
         :::|||::::::|||:::   ::::::   :::|||
    1547 ATGATCGCGACTGTGAGCCCGAGCGCGCTGAACTAC.............. 1582

      18 rValGluAsnSerSerGlnThrAlaSerGlnSerValSerAlaLysAspI 35
           |||::::::      |||        :::      |||:::||||
    1583 ....GAGGAGACGCTGAGCACGCTGCGGTACGCGTCGCGCGCGCGCGACA 1628

      35 leLeuAsnSerIleGlyIleSer............SerSerLysValSer 47
         ||:::|||         ::::::             :::   ::::::
    1629 TTGTGAATGTTGCGCAGGTGAACGAGGACCCGCGCGCACGGCGGATCCGC 1678

      48 AspLeu..............GlyLeuSerProThrLeuSerAlaProAl 59
         ::::|||              ::::::   ::::::::::::::::   ::
    1679 GAGCTGGAGGAGCAGATGGAGGACATGCGGCAGGCGATGGCTGGCGGCGA 1728

      59 aProGlyValLeuThrGlnThrProGlyThrIleThrSerPheLeuLysA 76
         :|||:::   :::::::::         ::::::   ::::::|||   :
    1729 CCCCGCGTACGTGTCTGAGCTGAAGAAGAAGCTT...GCGCTGCTGGAGT 1775

      76 laSerIleGlnAsnThrAspMetAsnGlnAspLeuAsnAlaLeuAlaAsn 92
         ::      |||:::          :::   ||||||:::||||||:::
    1776 CGGAGGCGCAGAAG......CGTGCGGCGGACCTGCAGGCGCTGGAGAGG 1819

      93 AsnValThrThrLysAlaAsnGluValValGlnThrGln...LeuArgGl 108

       :::         ::::::::::::   ||||||   :::   ||||||::
 1820 GAG.........CGGGAGCACAACCAGGTGCAGGAGCGGCTGCTGCGCGC 1860

  108 uGlnGlnAlaGluValGlyLysPhePheAspIleSerGlyMetSerSerS 125
      :   :::||||||                        |||   :::   |||
 1861 GACGGAGGCGGAG..................AAGAGCGAGCTGGAGTCGC 1892

  125 erAlaValAlaLeu.........LeuAlaAlaAla.............. 133
        |||:::||||||         ::::::|||:::
 1893 GTGCGGCTGCGCTGCAGGAGGAGATGACCGCGACTCGACGGCAGGCGGAC 1942

  134 ......AsnThrLeuMetLeuThrLeuAsnGlnAlaAspSerLysLeuSe 148
            :::::::|||   |||   |||::::::::::::::::::
 1943 AAGATGCAGGCGCTGAACCTGCGGCTGAAGGAAGAGCAGGCGCGCAAGGA 1992

  148 rGlyLysLeu......SerLeuValSerPheAspAlaAla.......... 159
                |||         ::::::      ||||||||
 1993 GCGCGAGCTGCTGAAAGAGATGGCGAAGAAGGACGCCGCGCTCTCGAAGG 2042

  160 .............................................Lys 160

 2043 TTCGGCGACGCAAAGACGCCGAGATAGCAAGCGAGCGCGAGAAGCTGGAG 2092

  161 ThrThrAlaSerSerMetMetArgGlu.................GlyMe 171
      :::|||::::::   :::   ||||||                  :::::
 2093 TCGACCGTGGCGCAGCTGGAGCGTGAGCAGCGCGAGCGCGAGGTGGCTCT 2142

  171 tAsnAlaLeuSerGlySerIleSer......GlnSerAlaLeuGlnLeuG 186
      ::::||||||                  |||   ||||||:::   :
 2143 GGACGCATTGCAGACGCACCAGAGAAAGCTGCAGGAAGCGCTCGAGAGCT 2192

  186 lyIleThrGlyValGlyAlaLys............LeuGluTyrLysGly 198
      ::         ::::::::::::            :::   :::
 2193 CTGAGCGGACAGCCGCGGAAAGGGACCAGCTGCTGCAGCAGCTAACAGAG 2242

  199 LeuGlnAsnGluArgGlyAlaLeuLysHisAsnAlaAlaLysIleAspLy 215
      ||||||:::||||||       |||   :::   ::::::       :::::
 2243 CTTCAGTCTGAGCGTACGCAGCTATCACAGGTTGTGACCGACCGCGAGCG 2292

  215 sLeuThrThrGluSerHisSerIleLys..................... 224
      :||||||   :::   :::   |||:::
 2293 GCTTACACGCGACTTGCAGCGTATTCAGTACGAGTACGGGGAAACCGAGC 2342

  224 ............................................... 224

 2343 TCGCGCGAGACGTGGCGCTGTGCGCCGCGCAGGAGATGGAGGCGCGCTAC 2392

  225 ......AsnValLeuAsnGlyGlnAsnSerValLysLeuGlyAlaGluGl 239
            :::|||::::::   |||:::   :::   |||::::::|||
 2393 CACGCTGCTGTGTTTCACCTGCAAACGCTCCTGGAGCTCGCAACCGAGTG 2442

  239 yValAspSerLeuLysSerLeuAsnMetLysLysThrGlyThrAspAlaT 256
        |||:::|||:::       ::::::            :::   :::|||:
 2443 GGAGGACGCACTCCGCGAGCGTGCGCTTGCAGAGCGTGACGAAGCCGCTG 2492

  256 hrLysAsnLeuAsnAspAlaThrLeuLysSer...AsnAlaGlyThrSer 271
      ::   :::|||::::::|||:::   :::|||   ||||||      |||
 2493 CAGCCGAACTTGATGCCGCAGCCTCTACTTCCCAAAACGCACGTGAAAGC 2542

  272 AlaThrGluSerLeuGlyIleLysAspSerAsnLysGlnIleSerProGl 288
      |||   |||   |||          ::   :::   :::   |||   ||
 2543 GCCTGCGAGCGGCTAACCAGCCTTGAGCAGCAGCTTCGCGAATCCGAGGA 2592

  288 uHisGlnAlaIleLeuSerLysArgLeuGluSerValGluSerAspIleA 305
      |:::   |||   |||:::   :::||||||:::::::::::::::
 2593 GCGCGCTGCGGAGCTGGCGAGCCAGCTGGAGGCCACTGCTGCTGCGAAGT 2642

  305 rgLeuGluGlnAsnThrMetAspMetThrArgIle.............. 316
            :::::::::      :::   ||||||
 2643 CGTCGGCGGAGCAGGACCGCGAGAACACGAGGGCCACGCTAGAGCAGCAG 2692

  317 ............AspAlaArgLysMetGlnMetThrGlyAspLeuIleMe 329
                  :::||||||      ::::::::::::::|||
```

```
DEFINITION  G.domesticus mRNA for ventricular myosin heavy chain.
ACCESSION   X59552
NID         g62995
VERSION     X59552.1  GI:62995
KEYWORDS    actin-activated ATPase activity; contractile protein; myosin heavy
            chain.
SOURCE      chicken.
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Vertebrata; Archosauria; Aves;
            Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 3263)
  AUTHORS   Stewart,A.F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-MAY-1991) A.F.R. Stewart, University of California,
            Dept of Anatomy, San Francisco, CA 94143-0452, USA
REFERENCE   2  (bases 1 to 3263)
  AUTHORS   Stewart,A.F., Camoretti-Mercado,B., Perlman,D., Gupta,M.,
            Jakovcic,S. and Zak,R.
  TITLE     Structural and phylogenetic analysis of the chicken ventricular
            myosin heavy chain rod
  JOURNAL   J. Mol. Evol. 33 (4), 357-366 (1991)
  MEDLINE   92130260
FEATURES             Location/Qualifiers
     source          1. .3263
                     /organism="Gallus gallus"
                     /strain="broiler breeders"
                     /db_xref="taxon:9031"
                     /dev_stage="adult"
                     /tissue_type="heart"
                     /clone="52.2, 52.4, 3B1, J, II"
                     /tissue_lib="lambda ZAP, chicken heart"
     CDS             <1. .3120
                     /codon_start=1
                     /product="myosin heavy chain"
                     /protein_id="CAA42130.1"
                     /db_xref="PID:g62996"
                     /db_xref="GI:62996"
                     /db_xref="SWISS-PROT:P29616"
                     /translation="TDAEERCDLLIKSKIQLEAKVKELTERAEDEEEMNSELTSKKRK
                     LEDEFALELKKDIDDLEITLAKVEKEKQATENKVKNLTEEMATLDENISKLTKEKKSL
                     QEAHQQVLDDLQAEEDKVNRLSKAKVKLEQQVDDLEGSLEQEKKVRMDLERAKRKLEG
                     DLKLTQESVMDLENDKLQMEEKLKKKEFEMSQLNSKIEDEQAIVMQLQKKIKELQARI
                     EELEEELEAERAARAKVEKQRSDLARELEELSERLEGAGGATAAQLEMNKKREAEFLK
                     LARDLEEATLHYEATAAALRKKHADRRGEMGEQLDNLQRVKQKLEKEKSELKMEVDDL
                     TANMEQTVKGKANAEKLWGTYEDHLNETKTKLDEMTRLMNDLTTQKTKLQSENGEFVR
                     QLEEKESLISQLSRGKTSFTQQIEELRRQLEEETKSKNALAHALQAARHDCDLLREQY
                     EEEQEAKAELQRALSKGNAEVAQWRTKYETDAIQRTEELEDAKKKLLARLQEAEEAIE
                     AANAKCSSLEKAKHRLQNEQEDMMIDLEKANSAAASLDKKQRGFDKIINDWKQKYEES
                     QAELEASQKEARSLSTELFKLKNAYEETLDHLETLKRENKNLQEEISDLTNQISEGNK
                     NLHEIEKVKKQVEQEKSEVQLALEEAEGALEHEESKTLRFQLELSQLKADFERKLAEK
                     DEEMQNIRRNQQRTIDSLQSTLDSEARSRNEAIRLKKKMEGDLNEMEIQLSHANRHAA
                     EATKSARGLQTQIKELQVQLDDLGHLNEDLKEQLAVSDRRNNLLQSELDELRALLDQT
                     ERARKLAEHELLEATERVNLLHTQNTSLINQKKKLEGDISQMQNEVEESIQECRNAEQ
                     KAKKAITDAAMMAEELKKEQDTSAHLERMKKNMEQTIKDLQKRLDEAEQIALKGGKKQ
                     IQKLESRVRELENELENELRRNSDAQKGARKFERRIKELTYQSEEDKKNLARMQDLID
                     KLQLKVKSYKHQAEEAEAQANLYLSKYRKQQHDLDDAEERAEIAESQVDKLRSKSRDI
                     GMKKVDAEE"
     polyA_signal    3220. .3225
     polyA_site      3240
BASE COUNT     1197 a    621 c    899 g    546 t
ORIGIN


alignment_scores:
            Quality:  110.00            Length:    353
              Ratio:  0.595              Gaps:     17
 Percent Similarity:  52.408   Percent Identity: 19.547

alignment_block:
 US-09-068-804-6 x GDVMHC   ..

 Align seg 1/1  to: GDVMHC  from: 1  to: 3263
```

```
 82 AspMetAsnGlnAspLeuAsnAlaLeuAlaAsnAsnValThr......Th 96
    :::::::::::::::::::::::::|||:::::|||:::::::      ||
238 AATCTGACAGAAGAAATGGCGACTCTTGATGAGAACATCAGCAAACTTAC 287

 96 rLysAlaAsn...........GluValValGlnThrGlnLeuArgGluG 109
    ||||::::::           |||:::   |||      |||   :::
288 TAAGGAGAAGAAGTCCTTGCAGGAAGCTCATCAGCAAGTTCTGGATGACC 337

109 lnGlnAlaGluValGlyLysPhePheAspIleSerGlyMetSerSerSer 125
      |||||||||   :::|||        ::::::   :::|||   :::
338 TTCAAGCAGAGGAAGACAAG.........GTCAACAGACTGAGCAAAGCT 378

126 AlaValAlaLeu..............................LeuAlaAl 132
      |||   |||                              |||:::
379 AAAGTGAAACTGGAACAGCAAGTGGATGATCTTGAGGGCTCGCTTGAGCA 428

132 aAlaAsnThrLeuMetLeuThrLeuAsnGlnAlaAspSerLysLeuSerG 149
     ::::::::::::   :::   |||::::::|||      ||||||   |
429 AGAGAAGAAAGTGAGGATGGATCTAGAAAGAGCAAAACGCAAACTGGAAG 478

149 lyLysLeuSerLeu..............ValSerPheAspAlaAlaLys 160
    ||  |||   |||              ::::::::::::::::::|||
479 GAGATTTGAAGCTGACCCAAGAGAGTGTCATGGACTTGGAGAATGATAAG 528

161 ThrThrAlaSerSerMetMetArgGlu.........GlyMetAsnAlaLe 174
                  :::::::::              |||:::   ||
529 TTGCAAATGGAAGAAAAGCTGAAAAAGAAAGAGTTTGAAATGAGCCAATT 578

174 uSerGlySerIleSerGlnSer..............AlaLeuGlnLeuG 186
    |::::::   |||   :::                  ||||||
579 GAATTCCAAGATAGAAGATGAACAAGCTATTGTAATGCAGCTGCAGAAGA 628

186 lyIleThrGlyValGlyAlaLysLeuGluTyrLysGlyLeuGlnAsnGlu 202
      |||:::   :::   |||:::::::|||         |||:::::|||
629 AGATAAAGGAACTACAGGCTCGTATAGAA......GAGCTGGAAGAGGAG 672

203 ArgGlyAlaLeuLysHisAsnAlaAlaLysIleAspLysLeuThrThrGl 219
        |||   :::   :::   ||||||:::::|||       :::::
673 CTGGAAGCAGAAAGAGCTGCTCGAGCCAAGGTGGAAAAGCAGAGATCAGA 722

219 uSerHisSerIleLysAsnValLeuAsnGlyGlnAsnSerValLysLeuG 236
    :              :::   |||          |||   :::|||
723 TTTGGCCCGAGAGCTGGAGGAATTA............AGTGAGCGGCTTG 760

236 lyAlaGluGlyValAspSerLeuLysSerLeuAsnMetLysLysThrGly 252
      :::::::|||   ::::::         |||:::|||:::|||:::
761 AAGGGGCTGGGGGTGCCACTGCTGCCCAGCTGGAGATGAACAAGAAACGT 810

253 ThrAspAla............ThrLysAsnLeuAsnAspAlaThrLeuLy 265
      ::::::           :::::::::|||::::::|||||||||
811 GAAGCTGAGTTCCTGAAGCTGGCGCGTGACCTCGAGGAGGCCACGCTGCA 860

265 sSerAsnAlaGlyThrSerAlaThrGluSerLeuGlyIleLysAspSerA 282
       :::|||   ::::::|||
861 CTATGAAGCCACAGCTGCTGCTCTG..................... 885

282 snLysGlnIleSerProGluHisGlnAlaIleLeuSerLysArgLeuGlu 298
      ::::::      ::::::::::::::::   ::::::   :::|||:::
886 ..AGGAAGAAGCATGCGGACAGGCGTGGCGAGATGGGGGAGCAGCTGGAC 933

299 SerValGlu...SerAspIleArgLeuGluGlnAsnThrMetAspMetTh 314
    :::::::::            :::||||||::::::::   ::::::
934 AACCTGCAGCGCGTCAAGCAGAAACTGGAAAAGGAGAAAAGCGAGCTG.. 981

314 rArgIleAspAlaArgLysMetGlnMetThrGlyAspLeuIleMetLysA 331
                  ||||||::::::   :::||||||:::   :::
982 ..............AAAATGGAAGTG...GATGATCTGACAGCCAACA 1012

331 snSerValThrValGlyGlyIleAlaGlyAlaSer.........GlyGln 344
          ||||||   |||   |||:::|||            |||
1013 TGGAGCAAACGGTTAAGGGAAAAGCAAATGCAGAAAAACTTTGGGGCACT 1062

345 TyrAlaAlaThrGlnGluArgSerGluGlnGlnIleSerGlnValAsnAs 361
```

```
                     STNNVGDNAKENGGSNKSDKESDSDKDTDGKDGTEYEGSFKANLFEVLLNYDAELNLN
                     PFKLFFTTDIFMFFFQQDHKYSEELREITRNVTTGNDLEDEEPLKAIQTISELLTTSL
                     TAADIRIPISYLTFLIYWLFGDFKATNDFLSDKSVIKSLLSFSYQIQDEDVTIKCLVT
                     MLLGVAYEFSSKESPFPRKEYFEFITKTLGKDNYASRIKQFKKDSYFSKVDMNEDSIL
                     TPELDETGLPKVYFSTYFIQLFNENIYRIRTALSHDPDEEPINKISFEEVEKLQRQCT
                     KLKGEITSLQTETESTHENLTEKLIALTNEHKELDEKYQILNSSHSSLKENFSILETE
                     LKNVRDSLDEMTQLRDVLETKDKENQTALLEYKSTIHKQEDSIKTLEKGLETILSQKK
                     KAEDGINKMGKDLFALSREMQAVEENCKNLQKEKDKSNVNHQKETKSLKEDIAAKITE
                     IKAINENLEEMKIQCNNLSKEKEHISKELVEYKSRFQSHDNLVAKLTEKLKSLANNYK
                     DMQAENESLIKAVEESKNESSIQLSNLQNKIDSMSQEKENFQIERGSIEKNIEQLKKT
                     ISDLEQTKEEIISKSDSSKDEYESQISLLKEKLETATTANDENVNKISELTKTREELE
                     AELAAYKNLKNELETKLETSEKALKEVKENEEHLKEEKIQLEKEATETKQQLNSLRAN
                     LESLEKEHEDLAAQLKKYEEQIANKERQYNEEISQLNDEITSTQQENESIKKKNDELE
                     GEVKAMKSTSEEQSNLKKSEIDALNLQIKELKKKNETNEASLLESIKSVESETVKIKE
                     LQDECNFKEKEVSELEDKLKASEDKNSKYLELQKESEKIKEELDAKTTELKIQLEKIT
                     NLSKAKEKSESELSRLKKTSSEERKNAEEQLEKLKNEIQIKNQAFEKERKLLNEGSST
                     ITQEYSEKINTLEDELIRLQNENELKAKEIDNTRSELEKVSLSNDELLEEKQNTIKSL
                     QDEILSYKDKITRNDEKLLSIERDNKRDLESLKEQLRAAQESKAKVEEGLKKLEEESS
                     KEKAELEKSKEMMKKLESTIESNETELKSSMETIRKSDEKLEQSKKSAEEDIKNLQHE
                     KSDLISRINESEKDIEELKSKLRIEAKSGSELETVKQELNNAQEKIRINAEENTVLKS
                     KLEDIERELKDKQAEIKSNQEEKELLTSRLKELEQELDSTQQKAQKSEEERRAEVRKF
                     QVEKSQLDEKAMLLETKYNDLVNKEQAWKRDEDTVKKTTDSQRQEIEKLAKELDNLKA
                     ENSKLKEANEDRSEIDDLMLLVTDLDEKNAKYRSKLKDLGVEISSDEEDDEEDDEEDE
                     EEGQVA"
     misc_feature    2265. .2275
                     /gene="USO1"
                     /note="ARS-consensus"
                     /db_xref="SGD:S0002216"
BASE COUNT      2503 a   1075 c   1242 g   1660 t
ORIGIN


alignment_scores:
            Quality:  115.00              Length:    411
              Ratio:  0.502                Gaps:     20
 Percent Similarity:  55.718   Percent Identity: 20.195

alignment_block:
 US-09-068-804-6 x SCYDL058W   ..

 Align seg 1/1  to: SCYDL058W  from: 1  to: 6480

      79 GlnAsnThrAspMetAsnGlnAspLeuAsnAlaLeuAlaAsnAsnValTh 95
         :::::|||:::   :::|||:::||||||:::|||   :::|||::: 
    4105 GAGGCCACTGAAACCAAACAGCAACTCAACAGCCTACGTGCCAATTTAGA 4154

      95 rThr......LysAlaAsnGluValValGlnThrGlnLeuArg..... 107
          :::      |||:::::|||   :::   :::|||||:::       
    4155 ATCATTAGAAAAAGAGCATGAAGATTTAGCAGCTCAGTTGAAGAAGTACG 4204

     108 ..GluGlnGlnAlaGluValGlyLysPhePhe......AspIleSerGly 121
           ||||||   |||:::      :::   :::      :::||||||    
    4205 AGGAGCAAATTGCCAACAAGGAAAGGCAATATAACGAGGAAATATCTCAA 4254

     122 MetSerSerSerAlaValAlaLeuLeuAlaAlaAlaAsnThrLeuMetLe 138
         :::::::::      ::::::          ::::::::::::::::::   
    4255 TTAAATGATGAGATTACTTCCACTCAACAAGAAAATGAATCAATTAAGAA 4304

     138 uThrLeuAsnGlnAlaAspSerLysLeuSerGlyLysLeuSerLeuValS 155
          :::   ::::::   ::::::   :::                        
    4305 AAAAAATGACGAACTGGAGGGCGAAGTT..................... 4332

     155 erPheAspAlaAlaLysThrThrAla.........SerSerMetMetArg 168
               |||   |||:::|||:::          |||:::::::::::::   
    4333 .....AAAGCAATGAAGAGCACTTCAGAGGAACAATCAAATTTGAAAAAG 4377

     169 GluGlyMetAsnAlaLeuSerGlySerIle..................... 178
             ::::::|||||:::      |||                        
    4378 TCAGAGATTGATGCCTTAAATTTGCAAATTAAAGAGCTGAAAAAGAAGAA 4427

     179 .......SerGlnSerAlaLeuGlnLeuGlyIleThrGlyVal...GlyA 192
              :::::::::::|||      :::|||::::::|||   ::::    
    4428 CGAAACGAACGAAGCCAGTTTATTGGAGTCAATCAAAAGTGTAGAATCAG 4477

     192 laLysLeuGluTyrLysGlyLeuGlnAsnGlu.................. 202
         ::::::::      |||   ||||||:::|||                  
    4478 AAACGGTAAAAATTAAAGAATTACAAGACGAATGTAACTTTAAAGAAAAG 4527

     203 ..............ArgGlyAlaLeuLysHisAsnAlaAlaLysIleAs 214
                         :::   ||||||   ::::::::::|||   ::  
    4528 GAAGTTAGTGAATTGGAAGACAAACTCAAGGCCTCAGAAGATAAGAACTC 4577

     214 pLys.........LeuThrThrGluSerHisSerIleLysAsnValLeuA 228
         :|||      |||   :::|||||||:::   |||||||:::   |||:  
    4578 TAAATATTTGGAGCTACAGAAGGAATCGGAGAAGATAAAAGAAGAACTAG 4627

     228 snGlyGlnAsnSer...ValLysLeuGlyAlaGluGlyValAspSerLeu 243
         ::::::::::::::   :::|||:::      |||   :::   ::::|||
    4628 ATGCTAAAACGACTGAACTAAAAATTCAATTGGAAAAGATCACTAATTTA 4677

     244 LysSerLeuAsn..........................MetLysLysTh 251
                 :::                           :::||||||||  
    4678 TCTAAGGCCAAGGAGAAATCAGAATCTGAGTTGTCTAGATTGAAGAAGAC 4727

     251 rGlyThrAspAlaThrLysAsnLeuAsnAsp......AlaThrLeuLysS 266
         |:::::::::::::   ||||||   ::::::      ::::::|||||:  
    4728 ATCTTCTGAGGAAAGAAAAAATGCAGAGGAGCAATTGGAAAAATTGAAAA 4777

     266 er............AsnAlaGlyThrSerAlaThrGluSerLeuGlyIle 278
         ::              ::::::   :::   ::::::|||         :::  
    4778 ATGAAATACAAATCAAAAATCAGGCCTTCGAAAAGGAAAGAAAGCTACTT 4827

     279 LysAspSerAsnLysGlnIleSerProGluHisGlnAlaIleLeuSerLy 295
         ::::::::::::      |||::::::|||:::            |||    
    4828 AATGAAGGATCTTCAACAATTACACAAGAATAT............TCTGA 4865

     295 sArgLeuGluSerValGluSerAsp...IleArgLeuGluGlnAsnThr. 310
          ::::::::::::::::|||::::::   ||||||||||:::::::::::  
    4866 AAAGATCAATACTTTGGAAGATGAATTAATTAGGCTACAAAATGAAAACG 4915

     311 .................MetAspMetThrArgIleAspAlaArgLysMet 321
                           :::|||   ||||||   :::      |||:::  
    4916 AATTAAAAGCCAAAGAGATCGATAATACGAGGAGTGAACTAGAGAAGGTT 4965

     322 GlnMetThrGlyAsp...LeuIleMetLysAsnSerValThrValGlyGl 337
            :::::::::|||   |||:::      :::      |||:::   ::  
    4966 AGTTTAAGCAATGATGAGCTTTTAGAAGAGAAACAAAACACAATAAAAAG 5015

     337 yIleAlaGlyAlaSerGlyGlnTyrAlaAlaThrGlnGluArgSerGluG 354
         ::::   ::::::         |||   ::::::      |||:::::::  
    5016 CTTACAAGATGAAATTCTTTCATATAAAGACAAAATAACTAGAAATGACG 5065

     354 lnGln.........IleSerGlnValAsnAsnArgValAlaSerThrAla 367
         :::::         |||   :::   |||:::|||        :::      
    5066 AAAAGCTTCTATCAATAGAGCGAGATAATAAACGAGACCTGGAGAGCTTG 5115

     368 SerAspGluAlaArg................................... 372
         ::::::   |||
    5116 AAAGAACAACTCCGGGCTGCACAAGAATCCAAAGCTAAAGTTGAAGAAGG 5165

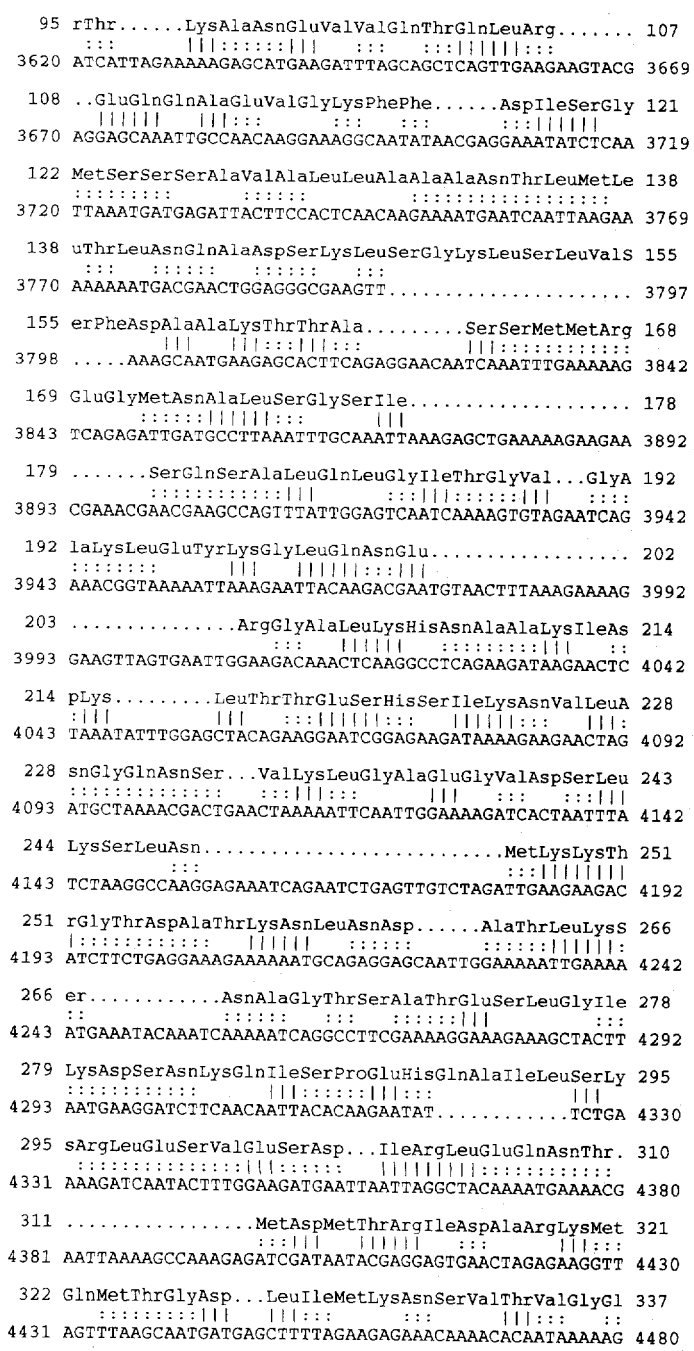
     373 ..................GluSerSerArgLysSerThrSerLeu.... 381
                           |||||||||:::      :::   |||      
    5166 GCTGAAAAAACTTGAGGAAGAATCTTCAAAAGAGAAAGCAGAGCTTGAAA 5215

     382 ..IleGlnGluMetLeuLysThrMetGluSerIleAsnGlnSerLysAla 397
           :::|||||:::|||::::::|||||:::   :::|||::::::      
    5216 AGTCTAAAGAAATGATGAAAAAGCTAGAGAGTACGATTGAGAGCAATGAA 5265

     398 SerAlaLeuAlaAlaIleAlaGlyAsnIleArg 408
         ::::::|||   :::         :::||||||   
    5266 ACGGAGTTGAAATCCTCAATGGAAACCATAAGG 5298

seq_name: gb_ov:GDVMHC

seq_documentation_block:
LOCUS       GDVMHC       3263 bp    mRNA            VRT       20-JUL-1995
```




```
seq_name: gb_pl1:SCYDL058W

seq_documentation_block:
LOCUS       SCYDL058W    6480 bp    DNA             PLN       11-AUG-1997
DEFINITION  S.cerevisiae chromosome IV reading frame ORF YDL058w.
ACCESSION   Z74106 Z71256
NID         g1431058
VERSION     Z74106.1  GI:1431058
KEYWORDS    .
SOURCE      baker's yeast.
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
            Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 6480)
  AUTHORS   Bloecker,H. and Brandt,P.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 6480)
  AUTHORS   MIPS.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUL-1996) Data collected by MIPS on behalf of the
            European yeast chromosome IV sequencing project. MIPS at the
            Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152
            Martinsried, FRG; E-mail: Mewes@mips.embnet.org
FEATURES             Location/Qualifiers
     source          1. .6480
                     /organism="Saccharomyces cerevisiae"
                     /db_xref="taxon:4932"
                     /chromosome="IV"
     gene            712. .6084
                     /gene="USO1"
                     /db_xref="SGD:S0002216"
     CDS             712. .6084
                     /gene="USO1"
                     /note="ORF YDL058w"
                     /codon_start=1
                     /db_xref="SGD:S0002216"
                     /protein_id="CAA98621.1"
                     /db_xref="PID:e253003"
                     /db_xref="PID:g1431059"
                     /db_xref="GI:1431059"
                     /db_xref="SPTREMBL:Q07380"
                     /translation="MDIIQGLIQQPKIQSVDETIPTLCDRVENSTLISDRRSAVLGLK
                     AFSRQYRESVIASGLKPLLNTLKRDYMDEDSVKAILETILILFIRGDGHDDLTRGWIS
                     QQSRLQNGKYPSPLVMKQEKEQVDQFSLWIADALTQSEDLIHLLVEFWEIDNFHIRLY
                     TIQLLEAVMATRPLKARSALISLPTSISTMVSLLDDMHEPIRDEAILLLMAVVNDSPH
                     VQKLVAFENIFERLFSIIEEEGGLRGSLVVNDCLSLINNILKYNTSNQTLFLETGNLP
                     KLAHLLSEPISQDEVFFWNDQRIVNINTALDIVSLTVEPGNTVTTKHQNALLDSSVLM
                     VVLRLAFFHNIPKKVRPVALLTAANMVRSNEHAQLEFSKIDVPYFDPSLPVNSTANGG
                     PIKLIPVVSILINWMLYANSVHTFDTRVACSRLLKAYFMDNFDLQRDFLLKQVQLCNN
```

```
 2403 TCTAAGGCCAAGGAGAAATCAGAATCTGAGTTGTCTAGATTGAAGAAGAC 2452

  251 rGlyThrAspAlaThrLysAsnLeuAsnAsp......AlaThrLeuLysS 266
      |::::::::::::   ||||||   ::::::      ::::::||||||:
 2453 ATCTTCTGAGGAAAGAAAAAATGCAGAGGAGCAATTGGAAAAATTGAAAA 2502

  266 er............AsnAlaGlyThrSerAlaThrGluSerLeuGlyIle 278
      ::           ::::::   :::   :::::::|||         :::
 2503 ATGAAATACAAATCAAAAATCAGGCCTTCGAAAAGGAAAGAAAGCTACTT 2552

  279 LysAspSerAsnLysGlnIleSerProGluHisGlnAlaIleLeuSerLy 295
      ::::::::::::      |||:::::::|||:::             |||
 2553 AATGAAGGATCTTCAACAATTACACAAGAATAT............TCTGA 2590

  295 sArgLeuGluSerValGluSerAsp...IleArgLeuGluGlnAsnThr. 310
       ::::::::::::::::|||::::::   ||||||||||:::::::::::
 2591 AAAGATCAATACTTTGGAAGATGAATTAATTAGGCTACAAAATGAAAACG 2640

  311 .................MetAspMetThrArgIleAspAlaArgLysMet 321
                      :::|||   ||||||   :::      |||:::
 2641 AATTAAAAGCCAAAGAGATCGATAATACGAGGAGTGAACTAGAGAAGGTT 2690

  322 GlnMetThrGlyAsp...LeuIleMetLysAsnSerValThrValGlyGl 337
      ::::::::::|||   |||:::      :::      |||:::   ::
 2691 AGTTTAAGCAATGATGAGCTTTTAGAAGAGAAACAAAACACAATAAAAAG 2740

  337 yIleAlaGlyAlaSerGlyGlnTyrAlaAlaThrGlnGluArgSerGluG 354
      ::::   ::::::         |||   ::::::      |||:::::::
 2741 CTTACAAGATGAAATTCTTTCATATAAAGACAAAATAACTAGAAATGACG 2790

  354 lnGln.........IleSerGlnValAsnAsnArgValAlaSerThrAla 367
      :::::         |||   :::   |||:::|||         :::
 2791 AAAAGCTTCTATCAATAGAGCGAGATAATAAACGAGACCTGGAGAGCTTG 2840

  368 SerAspGluAlaArg................................. 372
      ::::::   |||
 2841 AAAGAACAACTCCGGGCTGCACAAGAATCCAAAGCTAAAGTTGAAGAAGG 2890

  373 ...................GluSerSerArgLysSerThrSerLeu.... 381
                        |||||||||:::      :::   |||
 2891 GCTGAAAAAACTTGAGGAAGAATCTTCAAAAGAGAAAGCAGAGCTTGAAA 2940

  382 ..IleGlnGluMetLeuLysThrMetGluSerIleAsnGlnSerLysAla 397
        :::|||||:::|||::::::|||||:::   :::|||:::::::
 2941 AGTCTAAAGAAATGATGAAAAAGCTAGAGAGTACGATTGAGAGCAATGAA 2990

  398 SerAlaLeuAlaAlaIleAlaGlyAsnIleArg 408
      ::::::|||   :::         :::|||||||
 2991 ACGGAGTTGAAATCCTCAATGGAAACCATAAGG 3023
```

```
seq_name: gb_pl1:SCUSO1

seq_documentation_block:
LOCUS       SCUSO1       5877 bp    DNA             PLN       27-MAY-1991
DEFINITION  S.cerevisiae gene for Uso1 protein, involved in intracellular
            protein transport.
ACCESSION   X54378
NID         g4777
VERSION     X54378.1  GI:4777
KEYWORDS    protein transport; USO1 gene.
SOURCE      baker's yeast.
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
            Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 5877)
  AUTHORS   Nakajima,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-AUG-1990) Nakajima H., Dept of Bioengineering, Tokyo
            Institute of Technology, 2-12-1 O-okayama Meguro-ku, Tokyo 152,
            Japan
REFERENCE   2  (bases 1 to 5877)
  AUTHORS   Nakajima,H., Hirata,A., Ogawa,Y., Yonehara,T., Yoda,K. and
            Yamasaki,M.
  TITLE     A cytoskeleton-related gene, uso1, is required for intracellular
            protein transport in Saccharomyces cerevisiae
  JOURNAL   J. Cell Biol. 113 (2), 245-260 (1991)
  MEDLINE   91185402
FEATURES             Location/Qualifiers
     source          1. .5877
                     /organism="Saccharomyces cerevisiae"
                     /strain="X2180-1A"
                     /db_xref="taxon:4932"
                     /chromosome="IV"
     TATA_signal     155. .162
                     /note="potential"
     gene            177. .5549
                     /gene="USO1"
                     /db_xref="SGD:S0002216"
     CDS             177. .5549
                     /gene="USO1"
                     /codon_start=1
                     /db_xref="SGD:S0002216"
                     /product="Uso1 protein"
                     /protein_id="CAA38253.1"
                     /db_xref="PID:g4778"
                     /db_xref="GI:4778"
                     /db_xref="SWISS-PROT:P25386"
                     /translation="MDIIQGLIQQPKIQSVDETIPTLCDRVENSTLISDRRSAVLGLK
                     AFSRQYRESVIASGLKPLLNTLKRDYMDEDSVKAILETILILFIRGDGHDDLTRGWIS
                     QQSRLQNGKYPSPLVMKQEKEQVDQFSLWIADALTQSEDLIHLLVEFWEIDNFHIRLY
                     TIQLLEAVMATRPLKARSALISLPTSISTMVSLLDDMHEPIRDEAILLLMAVVNDSPH
                     VQKLVAFENIFERLFSIIEEEGGLRGSLVVNDCLSLINNILKYNISNQTLFLETGNLP
                     KLAHLLSEPISQDEVFFWNDQRIVNINTALDIVSLTVEPGNTVTTKHQNALLDSSVLM
                     VVLRLAFFHNIPKKVRPVALLTAANMVRSNEHAQLEFSKIDVPYFDPSLPVNSTATAG
                     PIKLIPVVSILINWMLYANSVHTFDTRVACSRLLKAYFMDNFDLQRDFLLKQVQLCNN
                     STNNVGDNAKENGGSNKSDKESDSDKDTDGKDGTEYEGSFKANLFEVLLNYDAELNLN
                     PFKLFFTTDIFMFFFQQDHKYSEELREITRNVTTGNDLEDEEPLKAIQTISELLTTSL
                     TAADIRIPISYLTFLIYWLFGDFKATNDFLSDKSVIKSLLSFSYQIQDEDVTIKCLVT
                     MLLGVAYEFSSKESPFPRKEYFEFITKTLGKDNYASRIKQFKKDSYFSKVDMNEDSIL
                     TPELDETGLPKVYFSTYFIQLFNENIYRIRTALSHDPDEEPISKISFEEVEKLQRQCT
                     KLKGEITSLQTETESTHENLTEKLIALTNEHKELDEKYQILNSSHSSLKENFSILETE
                     LKNVRDSLDEMTQLRDVLETKDKENQTALLEYKSTIHKQEDSIKTLEKGLETILSQKK
                     KAEDGINKMGKDLFALSREMQAVEENCKNLQKEKDKSNVNHQKETKSLKEDIAAKITE
                     IKAINENLEEMKIQCNNLSKEKEHISKELVEYKSRFQSHDNLVAKLTEKLKSLANNYK
                     DMQAENESLIKAVEESKNESSIQLSNLQNKIDSMSQEKENFQIERGSIEKNIEQLKKT
                     ISDLEQTKEEIISKSDSSKDEYESQISLLKEKLETATTANDENVNKISELTKTREELE
                     AELAAYKNLKNELETKLETSEKALKEVKENEEHLKEEKIQLEKEATETKQQLNSLRAN
                     LESLEKEHEDLAAQLKKYEEQIANKERQYNEEISQLNDEITSTQQENESIKKKNDELE
                     GEVKAMKSTSEEQSNLKKSEIDALNLQIKELKKKNETNEASLLESIKSVESETVKIKE
                     LQDECNFKEKEVSELEDKLKASEDKNSKYLELQKESEKIKEELDAKTTELKIQLEKIT
                     NLSKAKEKSESELSRLKKTSSEERKNAEEQLEKLKNEIQIKNQAFEKERKLLNEGSST
                     ITQEYSEKINTLEDELIRLQNENELKAKEIDNTRSELEKVSLSNDELLEEKQNTIKSL
                     QDEILSYKDKITRNDEKLLSIERDNKRDLESLKEQLRAAQESKAKVEEGLKKLEEESS
                     KEKAELEKSKEMMKKLESTIESNETELKSSMETIRKSDEKLEQSKKSAEEDIKNLQHE
                     KSDLISRINESEKDIEELKSKLRIEAKSGSELETVKQELNNAQEKIRINAEENTVLKS
                     KLEDIERELKDKQAEIKSNQEEKELLTSRLKELEQELDSTQQKAQKSEEERRAEVRKF
                     QVEKSQLDEKAMLLETKYNDLVNKEQAWKRDEDTVKKTTDSQRQEIEKLAKELDNLKA
                     ENSKLKEANEDRSEIDDLMLLVTDLDEKNAKYRSKLKDLGVEISSDEEDDEEDDEEDE
                     EEGQVA"
BASE COUNT     2304 a    950 c   1144 g   1479 t
ORIGIN


alignment_scores:
            Quality:  115.00              Length:    411
              Ratio:  0.502                 Gaps:     20
 Percent Similarity:  55.718   Percent Identity: 20.195

alignment_block:
 US-09-068-804-6 x SCUSO1   ..

 Align seg 1/1  to: SCUSO1  from: 1  to: 5877

      79 GlnAsnThrAspMetAsnGlnAspLeuAsnAlaLeuAlaAsnAsnValTh 95
         :::::::|||:::   :::|||:::||||||:::|||   :::|||:::
    3570 GAGGCCACTGAAACCAAACAGCAACTCAACAGCCTACGTGCCAATTTAGA 3619
```

```
    1009 CAAACAGAAACTGAACTTAAGGAGAAGAAGTTGAGAGTTGAGGATGCTCT 1058

     392 eAsnGlnSerLysAlaSerAlaLeuAlaAlaIle 403
         :|||   :::|||||||:::::   :::::::|||
    1059 AAACGCAACCAAGGCTGCCGTTGAGGAAGGTATT 1092

seq_name: gb_pl1:SCYDL057W

seq_documentation_block:
LOCUS       SCYDL057W    5648 bp    DNA             PLN       11-AUG-1997
DEFINITION  S.cerevisiae chromosome IV reading frame ORF YDL057w.
ACCESSION   Z74105 Z71256
NID         g1431056
VERSION     Z74105.1  GI:1431056
KEYWORDS    .
SOURCE      baker's yeast.
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
            Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 5648)
  AUTHORS   Bloecker,H. and Brandt,P.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 5648)
  AUTHORS   MIPS.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUL-1996) Data collected by MIPS on behalf of the
            European yeast chromosome IV sequencing project. MIPS at the
            Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152
            Martinsried, FRG; E-mail: Mewes@mips.embnet.org
FEATURES             Location/Qualifiers
     source          1. .5648
                     /organism="Saccharomyces cerevisiae"
                     /db_xref="taxon:4932"
                     /chromosome="IV"
     gene            1. .3809
                     /gene="USO1"
     CDS             <1. .3809
                     /gene="USO1"
                     /note="ORF YDL058w"
                     /codon_start=3
                     /protein_id="CAA98620.1"
                     /db_xref="PID:e308403"
                     /db_xref="PID:g1903280"
                     /db_xref="GI:1903280"
                     /db_xref="SPTREMBL:Q07380"
                     /translation="FQQDHKYSEELREITRNVTTGNDLEDEEPLKAIQTISELLTTSL
                     TAADIRIPISYLTFLIYWLFGDFKATNDFLSDKSVIKSLLSFSYQIQDEDVTIKCLVT
                     MLLGVAYEFSSKESPFPRKEYFEFITKTLGKDNYASRIKQFKKDSYFSKVDMNEDSIL
                     TPELDETGLPKVYFSTYFIQLFNENIYRIRTALSHDPDEEPINKISFEEVEKLQRQCT
                     KLKGEITSLQTETESTHENLTEKLIALTNEHKELDEKYQILNSSHSSLKENFSILETE
                     LKNVRDSLDEMTQLRDVLETKDKENQTALLEYKSTIHKQEDSIKTLEKGLETILSQKK
                     KAEDGINKMGKDLFALSREMQAVEENCKNLQKEKDKSNVNHQKETKSLKEDIAAKITE
                     IKAINENLEEMKIQCNNLSKEKEHISKELVEYKSRFQSHDNLVAKLTEKLKSLANNYK
                     DMQAENESLIKAVEESKNESSIQLSNLQNKIDSMSQEKENFQIERGSIEKNIEQLKKT
                     ISDLEQTKEEIISKSDSSKDEYESQISLLKEKLETATTANDENVNKISELTKTREELE
                     AELAAYKNLKNELETKLETSEKALKEVKENEEHLKEEKIQLEKEATETKQQLNSLRAN
                     LESLEKEHEDLAAQLKKYEEQIANKERQYNEEISQLNDEITSTQQENESIKKKNDELE
                     GEVKAMKSTSEEQSNLKKSEIDALNLQIKELKKKNETNEASLLESIKSVESETVKIKE
                     LQDECNFKEKEVSELEDKLKASEDKNSKYLELQKESEKIKEELDAKTTELKIQLEKIT
                     NLSKAKEKSESELSRLKKTSSEERKNAEEQLEKLKNEIQIKNQAFEKERKLLNEGSST
                     ITQEYSEKINTLEDELIRLQNENELKAKEIDNTRSELEKVSLSNDELLEEKQNTIKSL
                     QDEILSYKDKITRNDEKLLSIERDNKRDLESLKEQLRAAQESKAKVEEGLKKLEEESS
                     KEKAELEKSKEMMKKLESTIESNETELKSSMETIRKSDEKLEQSKKSAEEDIKNLQHE
                     KSDLISRINESEKDIEELKSKLRIEAKSGSELETVKQELNNAQEKIRINAEENTVLKS
                     KLEDIERELKDKQAEIKSNQEEKELLTSRLKELEQELDSTQQKAQKSEEERRAEVRKF
                     QVEKSQLDEKAMLLETKYNDLVNKEQAWKRDEDTVKKTTDSQRQEIEKLAKELDNLKA
                     ENSKLKEANEDRSEIDDLMLLVTDLDEKNAKYRSKLKDLGVEISSDEEDDEEDDEEDE
                     EEGQVA"
     CDS             4206. .5192
                     /note="ORF YDL057w"
                     /codon_start=1
                     /protein_id="CAA98619.1"
                     /db_xref="PID:e253208"
                     /db_xref="PID:g1431057"
                     /db_xref="GI:1431057"
                     /db_xref="SPTREMBL:Q07379"
                     /translation="MEKKHVTVQIQSAPPSYIKLEANEKFVYITSTMNGLSYQIAAIV
                     SYPEKRNSSTANKEDGKLLCKENKLALLLHGSQSHKNAIYQTLLAKRLAEFGYWVLRI
                     DFRGQGDSSDNCDPGLGRTLAQDLEDLSTVYQTVSDRSLRVQLYKTSTISLDVVVAHS
                     RGSLAMFKFCLKLHAAESPLPSHLINCAGRYDGRGLIERCTRLHPHWQAEGGFWANGP
                     RNGEYKDFWIPLSETYSIAGVCVPEFATIPQTCSVMSCYGMCDHIVPISAASNYARLF
                     EGRHSLKLIENADHNYYGIEGDPNALGLPIRRGRVNYSPLVVDLIMEYLQDT"
BASE COUNT      2266 a    926 c   1105 g   1351 t
ORIGIN


alignment_scores:
           Quality:  115.00              Length:    411
             Ratio:  0.502                 Gaps:     20
 Percent Similarity:  55.718   Percent Identity: 20.195

alignment_block:
 US-09-068-804-6 x SCYDL057W   ..

 Align seg 1/1  to: SCYDL057W  from: 1  to: 5648

      79 GlnAsnThrAspMetAsnGlnAspLeuAsnAlaLeuAlaAsnAsnValTh 95
         :::::::|||:::   :::|||:::||||||:::|||   :::|||:::
    1830 GAGGCCACTGAAACCAAACAGCAACTCAACAGCCTACGTGCCAATTTAGA 1879

      95 rThr......LysAlaAsnGluValValGlnThrGlnLeuArg..... 107
          :::      |||:::::::|||   :::   ::::||||||:::
    1880 ATCATTAGAAAAAGAGCATGAAGATTTAGCAGCTCAGTTGAAGAAGTACG 1929

     108 ..GluGlnGlnAlaGluValGlyLysPhePhe......AspIleSerGly 121
           ||||||   |||:::      :::   :::      :::|||||||
    1930 AGGAGCAAATTGCCAACAAGGAAAGGCAATATAACGAGGAAATATCTCAA 1979

     122 MetSerSerSerAlaValAlaLeuLeuAlaAlaAlaAsnThrLeuMetLe 138
         :::::::::      ::::::         ::::::::::::::::::
    1980 TTAAATGATGAGATTACTTCCACTCAACAAGAAAATGAATCAATTAAGAA 2029

     138 uThrLeuAsnGlnAlaAspSerLysLeuSerGlyLysLeuSerLeuValS 155
          :::   ::::::   ::::::   :::
    2030 AAAAAATGACGAACTGGAGGGCGAAGTT..................... 2057

     155 erPheAspAlaAlaLysThrThrAla.........SerSerMetMetArg 168
              |||   |||:::|||:::         |||:::::::::::::
    2058 .....AAAGCAATGAAGAGCACTTCAGAGGAACAATCAAATTTGAAAAAG 2102

     169 GluGlyMetAsnAlaLeuSerGlySerIle..................... 178
             ::::::||||||:::      |||
    2103 TCAGAGATTGATGCCTTAAATTTGCAAATTAAAGAGCTGAAAAAGAAGAA 2152

     179 .......SerGlnSerAlaLeuGlnLeuGlyIleThrGlyVal...GlyA 192
             ::::::::::::|||      :::|||:::::::|||   ::::
    2153 CGAAACGAACGAAGCCAGTTTATTGGAGTCAATCAAAAGTGTAGAATCAG 2202

     192 laLysLeuGluTyrLysGlyLeuGlnAsnGlu.................. 202
         ::::::::      |||   ||||||:::|||
    2203 AAACGGTAAAAATTAAAGAATTACAAGACGAATGTAACTTTAAAGAAAAG 2252

     203 ..............ArgGlyAlaLeuLysHisAsnAlaAlaLysIleAs 214
                        :::   ||||||   ::::::::::|||   ::
    2253 GAAGTTAGTGAATTGGAAGACAAACTCAAGGCCTCAGAAGATAAGAACTC 2302

     214 pLys.........LeuThrThrGluSerHisSerIleLysAsnValLeuA 228
         :|||          |||   :::|||||||:::   |||||||:::   |||:
    2303 TAAATATTTGGAGCTACAGAAGGAATCGGAGAAGATAAAAGAAGAACTAG 2352

     228 snGlyGlnAsnSer...ValLysLeuGlyAlaGluGlyValAspSerLeu 243
         ::::::::::::::   :::|||:::       |||   :::   :::|||
    2353 ATGCTAAAACGACTGAACTAAAAATTCAATTGGAAAAGATCACTAATTTA 2402

     244 LysSerLeuAsn.........................MetLysLysTh 251
              :::                             :::||||||||
```

```
REFERENCE   2  (bases 1 to 1707)
  AUTHORS   Schmitz,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JAN-1996) Schmitz G., Universitaet Frankfurt,
            Botanisches Institut, Siesmayerstr. 70, Frankfurt, Deutschland,
            60323
FEATURES             Location/Qualifiers
     source          1. .1707
                     /organism="Secale cereale"
                     /cultivar="Petkus 'Halo'"
                     /db_xref="taxon:4550"
                     /dev_stage="Seedling"
                     /tissue_type="Leaf"
     mRNA            <1. .1707
     gene            1. .1500
                     /gene="cpn60"
     CDS             <1. .1500
                     /gene="cpn60"
                     /codon_start=1
                     /product="chaperonin"
                     /protein_id="CAA93139.1"
                     /db_xref="PID:e219955"
                     /db_xref="PID:g1167858"
                     /db_xref="GI:1167858"
                     /db_xref="SWISS-PROT:Q43831"
                     /translation="PQIVNDGVTVAREVELEDPVENIGAKLVRQAAAKTNDLAGDGTT
                     TSVVLAQGLIAEGVKVIAAGANPVQITRGIEKTAKALVLELKKMSKEVEDSELADVAA
                     VSAGNNYEIGNMIAEAMSKVGRKGVVTLEEGRSSENNLYVVEGMQFERGYISPYFVTD
                     SEKMTTEYENCKLLLVDKKITNARDLINVLEEAIRGQYPILIIAEDIEQEALATLVVN
                     KLRGSLKICAIKAPGFGERKTQYLDDIAILTGGTVIRDEVGLTLDKADNTVLGTAAKV
                     VLTKESTTIVGDGSTQEEVTKRVAQIKNLIEAAEQDYEKEKLNERIAKLAGGVAVIQV
                     GAQTETELKEKKLRVEDALNATKAAVEEGIVVGGGCTLLRLAAKVDAIKDTLENDEQK
                     VGAEIVRRALCYPLKLIAKNAGVNGSVVTEKVLSNDNFKFGYNAATGQYEDLMAAGII
                     DPTKVVRCCLEHAASVAKTFLTSDVVVVEIKEPEAAPLANPMDNSGFGY"
BASE COUNT      472 a    347 c    496 g    392 t
ORIGIN


alignment_scores:
            Quality:  108.50              Length:    428
              Ratio:  0.464                Gaps:     17
 Percent Similarity:  54.673   Percent Identity: 17.290

alignment_block:
 US-09-068-804-6 x SCCPN60   ..

 Align seg 1/1  to: SCCPN60  from: 1  to: 1707

      19 ValGluAsnSerSerGlnThrAlaSerGlnSerValSerAlaLysAspIl 35
         :::   |||::::::   |||:::::::::   |||         |||
       7 ATTGTCAATGATGGTGTTACAGTTGCAAGAGAGGTTGAGCTGGAGGATCC 56

      35 eLeuAsnSerIleGlyIle..............SerSerSerLysValS 47
         .:::::::::||||||               ::::::::::|||::::
      57 TGTTGAAAATATTGGAGCTAAGTTGGTTAGGCAAGCTGCAGCGAAGACCA 106

      47 erAspLeuGly..........................LeuSerProThr 54
         ::|||||||:::                        |||::::::
     107 ACGATTTAGCTGGAGACGGGACAACTACTTCTGTCGTCCTTGCTCAGGGG 156

      55 LeuSerAlaProAlaProGlyValLeuThrGlnThrProGlyThrIleTh 71
         |||   |||   :::     |||::::::     ::::::::::::
     157 CTCATTGCTGAGGGTGTTAAGGTTATTGCAGCTGGTGCTAACCCTGTTCA 206

      71 rSerPheLeuLysAlaSerIleGlnAsnThrAspMetAsnGlnAspLeuA 88
            ::::::   :::|||:::::::|||:::              :
     207 A......ATTACTCGTGGTATTGAGAAAACAGCG.............A 235

      88 snAlaLeuAlaAsnAsnValThrThrLysAlaAsnGluValValGlnThr 104
         ::||||||:::   ::::::::::::::::::::||||||   ::::::
     236 AAGCTCTAGTCCTTGAACTTAAGAAGATGTCAAAGGAGGTTGAAGACAGT 285

     105 GlnLeuArgGluGlnGlnAla...GluValGlyLysPhePheAspIleSe 120
         :::|||   :::      |||      :::|||:::   :::::::|||::
     286 GAACTTGCCGATGTTGCTGCGGTCAGTGCTGGCAACAACTACGAAATCGG 335

     120 rGlyMetSerSerSerAlaValAlaLeuLeuAlaAlaAlaAsnThrLeuM 137
         :::::|||   :::   |||::::::   :::
     336 TAACATGATAGCAGAGGCTATGAGCAAGGTC..................... 366

     137 etLeuThrLeuAsnGlnAlaAspSerLysLeuSerGlyLysLeuSerLeu 153
                                        |||:::   ::::::
     367 ................................GGACGGAAGGGAGTG 381

     154 ValSerPheAspAlaAlaLysThrThrAlaSerSerMet...MetArgGl 169
         |||::::::::::::::::::::::::::::::::::::   :::   ||
     382 GTTACCCTTGAAGAGGGGAGGAGTTCCGAGAACAACCTTTATGTTGTGGA 431

     169 uGlyMetAsnAlaLeuSerGlySerIleSerGlnSerAlaLeuGlnLeuG 186
         |||||||:::         |||   ||||||:::
     432 GGGAATGCAGTTTGAGCGTGGTTATATCTCTCCT.................. 465

     186 lyIleThrGlyValGlyAlaLysLeuGluTyrLysGlyLeuGlnAsnGlu 202
                                     |||         :::::|||
     466 ........................TATTTCGTTACTGACAGCGAG 486

     203 ArgGlyAlaLeuLysHisAsnAlaAlaLys.........IleAspLysLe 216
         :::   :::      :::::::::   |||         :::|||||||
     487 AAAATGACCACCGAGTACGAGAACTGCAAGCTGCTTTTGGTTGACAAGAA 536

     216 uThrThrGluSerHisSerIleLysAsnValLeuAsn............G 229
          :::|||:::::::::::::::   |||||||||:::           |
     537 AATCACCAATGCCCGGGATCTAATCAATGTTCTGGAGGAAGCCATCAGGG 586

     229 lyGlnAsnSerValLysLeuGlyAlaGluGlyVal......AspSerLeu 243
         |||||   ::::::   :::   ||||||::::::      :::::|||
     587 GTCAATACCCAATCCTGATCATTGCTGAGGATATCGAGCAGGAGGCTCTT 636

     244 LysSerLeuAsnMetLysLysThrGlyThrAspAlaThrLysAsnLeuAs 260
            :::|||   ::::::|||
     637 GCAACCCTTGTTGTCAACAAG............................ 657

     260 nAspAlaThrLeuLysSerAsnAlaGlyThrSerAlaThrGluSerLeuG 277
                 |||:::::::::      ::::::|||:::   :::   |
     658 ..........CTGAGAGGCTCTTTGAAAATTTGCGCTATCAAAGCCCCTG 697

     277 lyIleLysAspSerAsnLysGlnIleSerProGluHisGlnAlaIleLeu 293
         ||:::   :::   ::::::|||          ::::::   |||||||||
     698 GTTTTGGAGAGCGCAAGACCCAGTAC...CTGGATGACATTGCCATCCTC 744

     294 SerLysArgLeuGluSerValGluSerAspIleArgLeuGluGlnAsnTh 310
         :::                      :::   :::      ::::::
     745 ACC..................GGAGGAACTGTAATCAGAGACGAGGTTGG 776

     310 rMetAspMetThrArgIleAspAlaArgLysMetGlnMetThrGlyAspL 327
         :::   :::   :::   |||:::      :::      ::::::   :
     777 ACTCACACTTGACAAGGCAGATAACACAGTTCTAGGAACGGCTGCAAAGG 826

     327 euIleMetLysAsnSerValThrValGlyGlyIleAlaGlyAlaSerGly 343
         ::::::::::::::      |||:::      |||::::|||:::
     827 TTGTCCTTACAAAAGAGTCGACAACA......ATAGTTGGTGAT...... 864

     344 GlnTyrAlaAlaThrGlnGluArgSerGluGlnGlnIleSerGlnValAs 360
               ::::::|||||||||         :::::::::::::|||:::::
     865 ......GGCAGCACCCAGGAAGAAGTGACTAAGAGGGTTGCACAGATCAA 908

     360 nAsnArgValAlaSerThrAlaSerAspGluAlaArgGlu......... 373
         :|||   ::::::::::::::   |||   :::::::|||
     909 AAATCTCATTGAGGCAGCAGAGCAAGACTACGAGAAGGAAAAACTCAACG 958

     374 ..........................................SerSer 375
                                                  ::::::
     959 AGAGGATTGCCAAGCTCGCCGGTGGTGTTGCTGTTATTCAGGTGGGAGCA 1008

     376 ArgLysSerThrSerLeuIleGlnGluMetLeuLysThrMetGluSerIl 392
         ::::::   |||   |||   :::   :::|||:::::::   ::::::::
```

```
                    KYLELQKESEKIKEELDAKTTELKIQLEKVTNLSKAKEKSESELSRLKKTSSEERKNA
                    EEQLEKLKNEIQIKNQAFEKERKLLNEGSSTITQEYSEKINTLEDELIRLQNENELKA
                    KEIDNTRSELEKVSLSNDELLEEKQNTIKSLQDEILSYKDKITRNDEKLLSIERDSKR
                    DLESLKEQLRAAQESKAKVEEGLKKLEEESSKEKAELEKSKEMMKKLESTIESNETEL
                    KSSMETIRKSDEKLEQSKKSAEEDIKNLQHEKSDLISRINESEKDIEELKSKLRIEAK
                    SSSELETVKQELNNAQEKIRVNAEENTVLKSKLEDIERELKDKQAEIKSNQEEKELLT
                    SRLKELEQELDSTQQKAQKSEEESRAEVRKFQVEKSQLDEKAMLLETKYNDLVNKEQA
                    WKRDEDTVKKTTDSQRQEIEKLAKELDNLKAENSKLKEANEDRSEIDDLMLLVTDLDE
                    KNAKYRSKLKDLGVEISSDEEDDEEEEDDEEDDEEDEEEGQVA"
     polyA_site     3609. .3614
     polyA_site     3642. .3647
BASE COUNT     1630 a    541 c    741 g    816 t
ORIGIN


alignment_scores:
            Quality:  113.00              Length:    411
              Ratio:  0.493                Gaps:     20
 Percent Similarity:  55.718   Percent Identity: 19.951

alignment_block:
 US-09-068-804-6 x YSCINTANA   ..

 Align seg 1/1  to: YSCINTANA  from: 1  to: 3728

      79 GlnAsnThrAspMetAsnGlnAspLeuAsnAlaLeuAlaAsnAsnValTh 95
         ::::::|||:::  :::|||:::||||||::|||  :::|||:::
    1052 GAGGCCACTGAAACCAAACAGCAACTCAACAGCCTACGTGCCAATTTAGA 1101

      95 rThr......LysAlaAsnGluValValGlnThrGlnLeuArg....... 107
          :::     |||:::::::||   :::   :::|||||||:::
    1102 ATCATTGGAAAAAGAGCATGAAGATTTAGCAGCTCAGTTGAAGAAGTACG 1151

     108 ..GluGlnGlnAlaGluValGlyLysPhePhe......AspIleSerGly 121
           ||||||   |||:::      :::   :::      :::||||||
    1152 AGGAGCAAATTGCCAACAAGGAAAGGCAATATAACGAGGAAATATCTCAA 1201

     122 MetSerSerSerAlaValAlaLeuLeuAlaAlaAlaAsnThrLeuMetLe 138
         :::::::::      ::::::      ::::::::::::::::::
    1202 TTAAATGATGAGATTACTTCCACTCAACAAGAAAATGAATCAATTAAGAA 1251

     138 uThrLeuAsnGlnAlaAspSerLysLeuSerGlyLysLeuSerLeuValS 155
          :::   ::::::   ::::::   :::
    1252 AAAAAATGACGAACTGGAGGGCGAAGTT..................... 1279

     155 erPheAspAlaAlaLysThrThrAla.........SerSerMetMetArg 168
                |||   |||:::|||:::         |||:::::::::::::
    1280 .....AAAGCAATGAAGAGCACTTCAGAGGAACAATCAAATTTGAAAAAG 1324

     169 GluGlyMetAsnAlaLeuSerGlySerIle..................... 178
             ::::::|||||||:::      |||
    1325 TCAGAGATTGATGCCTTAAATTTGCAAATTAAAGAGCTGAAAAAGAAGAA 1374

     179 .......SerGlnSerAlaLeuGlnLeuGlyIleThrGlyVal...GlyA 192
                 :::::::::::::|||      :::|||:::::::::   ::::
    1375 CGAAACGAACGAAGCCAGTTTATTGGAGTCAATCAAAAGTATAGAATCAG 1424

     192 laLysLeuGluTyrLysGlyLeuGlnAsnGlu.................... 202
          ::::::::      |||   |||||||:::|||
    1425 AAACGGTAAAAATTAAAGAATTACAAGACGAATGTAACTTTAAAGAAAAG 1474

     203 ..............ArgGlyAlaLeuLysHisAsnAlaAlaLysIleAs 214
                        :::   ||||||   ::::::::::|||   ::
    1475 GAAGTTAGTGAATTGGAAGACAAACTCAAGGCCTCAGAAGATAAGAACTC 1524

     214 pLys.........LeuThrThrGluSerHisSerIleLysAsnValLeuA 228
          :|||          |||   :::|||||||:::   ||||||:::   |||:
    1525 TAAATACTTGGAGCTACAGAAGGAATCGGAGAAGATAAAAGAAGAACTAG 1574

     228 snGlyGlnAsnSer...ValLysLeuGlyAlaGluGlyValAspSerLeu 243
          ::::::::::::::   :::|||:::      |||   |||  :::|||
    1575 ATGCTAAAACGACTGAACTAAAAATTCAATTGGAAAAGGTCACTAATTTA 1624

      244 LysSerLeuAsn.........................MetLysLysTh 251
                ::::                          :::|||||||||
     1625 TCTAAGGCCAAGGAGAAATCAGAGTCTGAGTTGTCTAGATTGAAGAAGAC 1674

      251 rGlyThrAspAlaThrLysAsnLeuAsnAsp......AlaThrLeuLysS 266
          |::::::::::::   ||||||   ::::::      :::::::||||||:
     1675 ATCTTCTGAGGAAAGAAAAAATGCAGAGGAGCAATTGGAAAAATTGAAAA 1724

      266 er............AsnAlaGlyThrSerAlaThrGluSerLeuGlyIle 278
          ::            ::::::   :::   :::::::|||        :::
     1725 ATGAAATACAAATCAAAAATCAGGCCTTCGAAAAGGAAAGAAAGCTACTT 1774

      279 LysAspSerAsnLysGlnIleSerProGluHisGlnAlaIleLeuSerLy 295
          ::::::::::::      |||:::::::|||:::             |||
     1775 AATGAAGGATCTTCAACAATTACACAAGAATAT............TCTGA 1812

      295 sArgLeuGluSerValGluSerAsp...IleArgLeuGluGlnAsnThr. 310
           ::::::::::::::::|||::::::   ||||||||:::::::::::::
     1813 AAAGATCAATACTTTGGAAGATGAATTAATTAGGCTACAAAATGAAAACG 1862

      311 .................MetAspMetThrArgIleAspAlaArgLysMet 321
                          :::|||   ||||||   :::      |||:::
     1863 AATTAAAAGCCAAAGAGATCGATAATACGAGGAGTGAACTAGAGAAGGTT 1912

      322 GlnMetThrGlyAsp...LeuIleMetLysAsnSerValThrValGlyGl 337
          ::::::::::|||   |||:::      :::      |||:::   ::
     1913 AGTTTAAGCAATGATGAGCTTTTAGAAGAGAAACAAAACACAATAAAAAG 1962

      337 yIleAlaGlyAlaSerGlyGlnTyrAlaAlaThrGlnGluArgSerGluG 354
          ::::   ::::::         |||   ::::::      |||::::::::
     1963 CTTACAAGATGAAATTCTTTCATATAAAGACAAAATAACTAGAAATGACG 2012

      354 lnGln.........IleSerGlnValAsnAsnArgValAlaSerThrAla 367
          :::::         |||   :::   :::::::|||         :::
     2013 AAAAGCTTCTATCAATAGAGCGAGATAGTAAACGAGACCTGGAGAGCTTG 2062

      368 SerAspGluAlaArg................................... 372
          ::::::   |||
     2063 AAAGAACAGCTCCGGGCTGCACAAGAATCCAAAGCTAAAGTTGAAGAAGG 2112

      373 ...................GluSerSerArgLysSerThrSerLeu.... 381
                             |||||||||:::      :::   |||
     2113 GCTGAAAAAACTTGAGGAAGAATCTTCAAAAGAGAAAGCAGAGCTTGAAA 2162

      382 ..IleGlnGluMetLeuLysThrMetGluSerIleAsnGlnSerLysAla 397
            :::|||||||:::|||:::::::|||||:::   :::|||:::::
     2163 AGTCTAAAGAAATGATGAAAAAGCTAGAGAGTACGATTGAGAGCAATGAA 2212

      398 SerAlaLeuAlaAlaIleAlaGlyAsnIleArg 408
          ::::::|||   :::         :::||||||
     2213 ACGGAGTTGAAATCCTCAATGGAAACCATAAGG 2245

seq_name: gb_pl1:SCCPN60

seq_documentation_block:
LOCUS       SCCPN60      1707 bp    mRNA            PLN       27-JAN-1997
DEFINITION  S.cereale cv. Petkus 'Halo' encoding cpn60.
ACCESSION   Z68903
NID         g1167857
VERSION     Z68903.1  GI:1167857
KEYWORDS    cpn60 gene.
SOURCE      rye.
  ORGANISM  Secale cereale
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Secale.
REFERENCE   1  (bases 1 to 1707)
  AUTHORS   Schmitz,G., Schmidt,M. and Feierabend,J.
  TITLE     Comparison of the expression of a plastidic chaperonin 60 in
            different plant tissues and under photosynthetic and
            non-photosynthetic conditions
  JOURNAL   Planta 200 (3), 326-336 (1996)
  MEDLINE   97085078
```

```
BASE COUNT     2518 a   1369 c   1135 g   1847 t
ORIGIN


alignment_scores:
            Quality:  377.50              Length:    355
              Ratio:  1.593                Gaps:      4
 Percent Similarity:  66.761   Percent Identity: 31.549

alignment_block:
 US-09-068-804-6 x SDIPABCD   ..

 Align seq 1/1  to: SDIPABCD  from: 1  to: 6869

      55 LeuSerAlaProAlaProGlyValLeuThrGlnThrProGlyThrIleTh 71
    3358 CTTAATGTCGGTAAAAATCCCGTATTAACAACCACATTAAATGAT...GA 3404

      71 rSerPheLeuLysAlaSerIleGlnAsnThrAspMetAsnGlnAspLeuA 88
    3405 TCAACTTTTAAAGTTATCAGAGCAGGTTCAGCATGATTCAGAAATCATTG 3454

      88 snAlaLeuAlaAsnAsnValThrThrLysAlaAsnGluValValGlnThr 104
    3455 CTCGCCTTACTGACAAAAAGATGAAAGATCTTTCAGAGATGAGTCACACC 3504

     105 GlnLeuArgGluGlnGlnAlaGluValGlyLysPhePheAspIleSerGl 121
    3505 CTTACTCCAGAGAACACT.................CTGGATATTTCCAG 3536

     121 yMetSerSerSerAlaValAlaLeuLeuAlaAlaAlaAsnThrLeuMetL 138
    3537 TCTTTCTTCTAATGCTGTTTCTTTAATTATTAGTGTAGCCGTTCTACTTT 3586

     138 euThrLeuAsnGlnAlaAspSerLysLeuSerGlyLysLeuSerLeuVal 154
    3587 CTGCTCTCCGCACTGCAGAAACTAAATTGGGCTCTCAATTGTCATTGATT 3636

     155 SerPheAspAlaAlaLysThrThrAlaSerSerMetMetArgGluGlyMe 171
    3637 GCGTTCGATGCTACAAAATCAGCTGCAGAGAACATTGTTCGGCAAGGCCT 3686

     171 tAsnAlaLeuSerGlySerIleSerGlnSerAlaLeuGlnLeuGlyIleT 188
    3687 GGCAGCCCTATCATCAAGCATTACTGGAGCAGTCACACAAGTAGGTATAA 3736

     188 hrGlyValGlyAlaLysLeuGluTyrLysGlyLeuGlnAsnGluArgGly 204
    3737 CGGGTATCGGTGCCAAAAAAACGCATTCAGGGATTAGCGACCAAAAAGGA 3786

     205 AlaLeuLysHisAsnAlaAlaLysIleAspLysLeuThrThrGluSerHi 221
    3787 GCCTTAAGAAAGAACCTTGCCACTGCTCAATCTCTTGAAAAAGAGCTTGC 3836

     221 sSerIleLysAsnValLeuAsnGlyGlnAsnSerValLysLeuGlyAlaG 238
    3837 AGGTTCTAAATTAGGGTTAAATAAACAAATAGATACAAATATCACCTCAC 3886

     238 luGlyValAspSerLeuLysSerLeuAsnMetLysLysThrGlyThrAsp 254
    3887 CACAAACTAACTCT..................................... 3900

     255 AlaThrLysAsnLeuAsnAspAlaThrLeuLysSerAsnAlaGlyThrSe 271
    3901 AGCACAAAATTTTTAGGTAAAAATAAACTGGCGCCAGAT........... 3939

     271 rAlaThrGluSerLeuGlyIleLysAspSerAsnLysGlnIleSerProG 288
    3940 ...............................AATATATCCCTGTCAACTG 3958

     288 luHisGlnAlaIleLeuSerLysArgLeuGluSerValGluSerAspIle 304
    3959 AACATAAAACTTCTCTTAGTTCTCCCGATATTTCTTTGCAGGATAAAATT 4008

     305 ArgLeuGluGlnAsnThrMetAspMetThrArgIleAspAlaArgLysMe 321
    4009 GACACCCAGAGAAGAACTTACGAGCTCAATACCCTTTCTGCGCAGCAAAA 4058

     321 tGlnMetThrGlyAspLeuIleMetLysAsnSerValThrValGlyGlyI 338
    4059 ACAAAACATTGGCCGTGCAACAATGGAAACATCAGCCGTTGCTGGTAATA 4108

     338 leAlaGlyAlaSerGlyGlnTyrAlaAlaThrGlnGluArgSerGluGln 354
    4109 TATCCACATCAGGAGGGCGTTATGCATCTGCTCTTGAAGAAGAAGAACAA 4158

     355 GlnIleSerGlnValAsnAsnArgValAlaSerThrAlaSerAspGluAl 371
    4159 CTAATCAGTCAGGCCAGCAGTAAACAAGCAGAGGAAGCATCCCAAGTATC 4208

     371 aArgGluSerSerArgLysSerThrSerLeuIleGlnGluMetLeuLysT 388
    4209 TAAAGAAGCATCCCAAGCGACAAATCAATTAATACAAAAATTACTGAATA 4258

     388 hrMetGluSerIleAsnGlnSerLysAlaSerAlaLeuAlaAlaIleAla 404
    4259 TAATTGACAACATCAACCAATCAAGGAGTTCTACAGCCAGTCAGATTGCT 4308

     405 GlyAsnIleArgAla 409
    4309 GGTAACATTCGAGCT 4323

seq_name: gb_pl1:YSCINTANA

seq_documentation_block:
LOCUS       YSCINTANA    3728 bp    DNA             PLN       17-MAY-1996
DEFINITION  Saccharomyces cerevisiae integrin analogue gene, complete cds.
ACCESSION   L03188
NID         g677197
VERSION     L03188.1  GI:677197
KEYWORDS    .
SOURCE      Saccharomyces cerevisiae (library: lambda gt11) DNA.
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
            Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 3728)
  AUTHORS   Hostetter,M.K., Tao,N.J., Gale,C., Herman,D.J., McClellan,M.,
            Sharp,R.L. and Kendrick,K.E.
  TITLE     Antigenic and functional conservation of an integrin I-domain in
            Saccharomyces cerevisiae
  JOURNAL   Biochem Mol Med 55 (2), 122-130 (1995)
  MEDLINE   96039111
COMMENT     On Feb 22, 1995 this sequence version replaced gi:677196.
FEATURES             Location/Qualifiers
     source          1. .3728
                     /organism="Saccharomyces cerevisiae"
                     /db_xref="taxon:4932"
                     /tissue_lib="lambda gt11"
     CDS             1. .3049
                     /partial
                     /standard_name="Int-1"
                     /note="putative"
                     /codon_start=2
                     /function="integrin analogue"
                     /protein_id="AAB00143.1"
                     /db_xref="PID:g677198"
                     /db_xref="GI:677198"
                     /translation="NSSHSSLKENFSILETELKNVRDSLDEMTQLRDVLETKDKENQT
                     ALLEYKSTIHKQEDSIKTLEKELETILSQKKKAEDGINKMGKDLFALSREMQAVEENC
                     KNLQKEKDKSNVNHQKETKSLKEDIAAKITEIKAINENLEKMKIQCNNLSKEKEHISK
                     ELVEYKSRFQSHDNLVAKLTEKLKSLANNYKDMQAENESLIKAVEESKNESSIQLSNL
                     QNKIDSMSQEKENFQIERGSIEKNIEQLKKTISDLEQTKEEIISKSDSSKDEYESQIS
                     LLKEKLETATTANDENVNKISELTKTREELEAELAAYKNLKNELETKLETSEKALKEV
                     KENEEHLKEEKIQLEKEATETKQQLNSLRANLESLEKEHEDLAAQLKKYEEQIANKER
                     QYNEEISQLNDEITSTQQENESIKKKNDELEGEVKAMKSTSEEQSNLKKSEIDALNLQ
                     IKELKKKNETNEASLLESIKSIESETVKIKELQDECNFKEKEVSELEDKLKASEDKNS
```

```
    4893 TTCTAAATTAGGGTTAAATAAACAAATAGATACAAATATCACCTCACCAC 4942

     239 lyValAspSerLeuLysSerLeuAsnMetLysLysThrGlyThrAspAla 255
           ::::::|||                                    :::
    4943 AAACTAACTCT.................................AGC 4956

     256 ThrLysAsnLeuAsnAspAlaThrLeuLysSerAsnAlaGlyThrSerAl 272
         ||||||   |||:::   ::::::|||   ::::::
    4957 ACAAAATTTTTAGGTAAAAATAAACTGGCGCCAGAT.............. 4992

     272 aThrGluSerLeuGlyIleLysAspSerAsnLysGlnIleSerProGluH 289
                                  |||      :::|||:::||||
    4993 ..........................AATATATCCCTGTCAACTGAAC 5014

     289 isGlnAlaIleLeuSerLysArgLeuGluSerValGluSerAspIleArg 305
         ||::::::   ||||||           |||::::::::::   |||
    5015 ATAAAACTTCTCTTAGTTCTCCCGATATTTCTTTGCAGGATAAAATTGAC 5064

     306 LeuGluGlnAsnThrMetAspMetThrArgIleAspAlaArgLysMetGl 322
            ::::::   |||   :::::::::   ::::::|||::::::::::||
    5065 ACCCAGAGAAGAACTTACGAGCTCAATACCCTTTCTGCGCAGCAAAAACA 5114

     322 nMetThrGlyAspLeuIleMetLysAsnSerValThrValGlyGlyIleA 339
         |   :::|||      :::|||   :::|||::::::::::|||:::|||:
    5115 AAACATTGGCCGTGCAACAATGGAAACATCAGCCGTTGCTGGTAATATAT 5164

     339 laGlyAlaSerGlyGlnTyrAlaAlaThrGlnGluArgSerGluGlnGln 355
         ::   ::::::|||:::|||||::::::   |||      ||||||
    5165 CCACATCAGGAGGGCGTTATGCATCTGCTCTTGAAGAAGAAGAACAACTA 5214

     356 IleSerGlnValAsnAsnArgValAlaSerThrAlaSerAspGluAlaAr 372
         |||||||||:::::::::::::   |||      ||||||:::   :::::
    5215 ATCAGTCAGGCCAGCAGTAAACAAGCAGAGGAAGCATCCCAAGTATCTAA 5264

     372 gGluSerSerArgLysSerThrSerLeuIleGlnGluMetLeuLysThrM 389
         :|||:::|||:::   ::::::   |||||||||   :::|||:::::::
    5265 AGAAGCATCCCAAGCGACAAATCAATTAATACAAAAATTATTGAATATAA 5314

     389 etGluSerIleAsnGlnSerLysAlaSerAlaLeuAlaAlaIleAlaGly 405
         :::::|||||||||||||||||||:::|||:::   :::   |||||||||
    5315 TTGACAGCATCAACCAATCAAAGAATTCGACAGCCAGTCAGATTGCTGGT 5364

     406 AsnIleArgAla 409
         ||||||||||||
    5365 AACATTCGAGCT 5376

seq_name: gb_ba1:SDIPABCD

seq_documentation_block:
LOCUS       SDIPABCD     6869 bp    DNA             BCT       24-JUL-1991
DEFINITION  Shigella dysenteriae ipaBCD genes.
ACCESSION   X60777
NID         g46932
VERSION     X60777.1  GI:46932
KEYWORDS    ipaB gene; ipaC gene; ipaD gene; ippI gene.
SOURCE      Shigella dysenteriae.
  ORGANISM  Shigella dysenteriae
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Shigella.
REFERENCE   1  (bases 1 to 6869)
  AUTHORS   Palchaudhuri,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUN-1991) S. Palchaudhuri, Wayne State Univers.School
            of Medicine, Department of Immunology and, Microbiology, 540 E.
            Canfield, Detroit, MI 48201, USA
REFERENCE   2  (bases 1 to 6869)
  AUTHORS   Yao,R. and Palchaudhuri,S.
  TITLE     Nucleotide sequence of the ipaBCD structure genes of Shigella
            dysenteriae
  JOURNAL   Unpublished
FEATURES             Location/Qualifiers
     source          1. .6869
                     /organism="Shigella dysenteriae"
                     /strain="CG097"
                     /db_xref="taxon:622"
     CDS             317. .943
                     /note="24kD ORF"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA43188.1"
                     /db_xref="PID:g46933"
                     /db_xref="GI:46933"
                     /db_xref="SWISS-PROT:Q04314"
                     /translation="MQILNKILPQVEFAIPRPSFNSLSRNKLVKKILSVFNLKQRFPQ
                     KNFGCTVNINKIRDNVIDKIKDSNSGNQLFCWMSQERTTYVSSMINRSIDEMAIHNGV
                     VLTSDNKKNIFAAIEKKFPDIKLDEKSAQTSISHTALNEIASSGLRAKILKRYSSDMD
                     LFNTQMKDLTNLVSSSVYDKIFNESTKVLQIEISAEVLKAVYRQSNTN"
     CDS             1000. .1467
                     /note="IppI ORF"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA43189.1"
                     /db_xref="PID:g46934"
                     /db_xref="GI:46934"
                     /db_xref="SWISS-PROT:P18008"
                     /translation="MSLNIITENESISTAVIDAINSGATLKDINAIPDDMMDDIYSYAY
                     DFYNKGRIEEAEVFFRFLCIYDFYNVDYIMGLAAIYQIKEQFQQAADLYAVAFALGKN
                     DYTPVFHTGQCQLRLKAPLKAKECFELVIQHSNDEKLKIKAQSYLDAIQDIKE"
     CDS             1473. .3215
                     /note="IpaB ORF"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA43190.1"
                     /db_xref="PID:g46935"
                     /db_xref="GI:46935"
                     /db_xref="SWISS-PROT:Q03945"
                     /translation="MHNVNTTTTGLSLAKILASTELGDNTIQAGNDAANKLFSLTIAD
                     LTANKNINTTNAHSTSNILIPELKAPKSLNASSQLTLLIGNLIQILGEKSLTALTNKI
                     TAWKSQQQARQQKNLEFSDKINTLLSETEGLTRDYEKQINKLKNADSKIKDLENKINQ
                     IQTRLSELDPDSPEKKKLSREEIQLTIKKDAAVKDRTLIEQKTLSIHSKLTDKSMQLE
                     KEIDSFSAFSNTASAEQLSTQQKSLTGLASVTQLMATFIQLVGKNNEESLKNDLALFQ
                     SLQESRKTEMERKSDEYAAEVRKAEELNRVMGCVGKILGALLTIVSVVAAAFSGGASL
                     ALADVGLALMVTDAIVQAATGNSFMEQALNPIMKAVIEPLIKLLSDAFTKMLEGLGVD
                     SKKAKMIGSILGAIAGALVLVAAVVLVATVGKQAAAKLAENIGKIIGKTLTDLIPKFL
                     KNFSSQLDDLITNAVARLNKFLGAAGDEVISKQIISTHLNQAVLLGESVNSATQAGGS
                     VASAVFQNSASTNLADLTLSKYQVEQLSKYISEAIEKFGQLQEVIADLLASMSNSQAN
                     RTDVAKAILQQTTA"
     CDS             3178. .4326
                     /note="IpaC ORF"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA43191.1"
                     /db_xref="PID:g46936"
                     /db_xref="GI:46936"
                     /db_xref="SWISS-PROT:Q03946"
                     /translation="MLQKQFCNKLLLDTNKENVMEIQNTKSTQILYTDISTKQTQSSS
                     ETQKSQNYQQIAAHIPLNVGKNPVLTTTLNDDQLLKLSEQVQHDSEIIARLTDKKMKD
                     LSEMSHTLTPENTLDISSLSSNAVSLIISVAVLLSALRTAETKLGSQLSLIAFDATKS
                     AAENIVRQGLAALSSSITGAVTQVGITGIGAKKTHSGISDQKGALRKNLATAQSLEKE
                     LAGSKLGLNKQIDTNITSPQTNSSTKFLGKNKLAPDNISLSTEHKTSLSSPDISLQDK
                     IDTQRRTYELNTLSAQQKQNIGRATMETSAVAGNISTSGGRYASALEEEEQLISQASS
                     KQAEEASQVSKEASQATNQLIQKLLNIIDNINQSRSSTASQIAGNIRA"
     CDS             4377. .5375
                     /note="IpaD ORF"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA43192.1"
                     /db_xref="PID:g46937"
                     /db_xref="GI:46937"
                     /db_xref="SWISS-PROT:Q03947"
                     /translation="MNITTLTNSISTSSFSPNNINGSSTETVNSDIKTTTSSHPVSSL
                     TMLNDTLHNIRTTNQALKKELSQKTLTKTSLEEIALHSSQISMDVNKSAQLLNILSKT
                     EYPINKDARELLHSAPKEAELDGYEMISHRELWAKIANSINDINEQYLKVYEHAVSSY
                     TQMYQEFSAVLSSLAGWISPGGNDGNSVKLQVKSLKDALTTLKKNYEDKPLYPATNTV
                     SEQEANKWLTELGGTIGTVSAKNGGYVVSINMTPIYNMLNRLDNLGGNGEVVLDNAKY
                     QAWNAGFSAEDETMKNNLQTLVQKYSNANSIFDNLVKVLSSTISSCTDTDKLFLHF"
```

```
                     /db_xref="PID:g47053"
                     /db_xref="GI:47053"
                     /db_xref="SWISS-PROT:P33548"
                     /translation="MQILNKILPQVEFAIPRPSFDSLSRNKLVKKILSVFNLKQRFPQ
                     KNFGCPVNINKIRDSVIDKIKDSNSGNQLFCWMSQERTTYVSSMINRSIDEMAIHNGV
                     VLTSDNKKNIFAAIEKKFPDIKLDEKSAQTSISHTALNEIASSGLRAKILKRYSSDMD
                     LFNTQMKDLTNLVSSSVYDKIFNESTKVLQIEISAEVLKAVYRQSNTN"
RBS                  2043. .2045
                     /note="put. ribosome binding site"
CDS                  2053. .2520
                     /note="18k gene (AA 1-155)"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA33380.1"
                     /db_xref="PID:g47054"
                     /db_xref="GI:47054"
                     /db_xref="SWISS-PROT:P18008"
                     /translation="MSLNITENESISTAVIDAINSGATLKDINAIPDDMMDDIYSYAY
                     DFYNKGRIEEAEVFFRFLCIYDFYNVDYIMGLAAIYQIKEQFQQAADLYAVAFALGKN
                     DYTPVFHTGQCQLRLKAPLKAKECFELVIQHSNDEKLKIKAQSYLDAIQDIKE"
RBS                  2513. .2515
                     /note="put. ribosome binding site"
CDS                  2526. .4268
                     /note="ipaB gene (AA 1-580)"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA33381.1"
                     /db_xref="PID:g47055"
                     /db_xref="GI:47055"
                     /db_xref="SWISS-PROT:P18011"
                     /translation="MHNVSTTTTGFPLAKILASTELGDNTIQAANDAANKLFSLTIAD
                     LTANQNINTTNAHSTSNILIPELKAPKSLNASSQLTLLIGNLIQILGEKSLTALTNKI
                     TAWKSQQQARQQKNLEFSDKINTLLSETEGLTRDYEKQINKLKNADSKIKDLENKINQ
                     IQTRLSELDPESPEKKKLSREEIQLTIKKDAAVKDRTLIEQKTLSIHSKLTDKSMQLE
                     KEIDSFSAFSNTASAEQLSTQQKSLTGLASVTQLMATFIQLVGKNNEESLKNDLALFQ
                     SLQESRKTEMERKSDEYAAEVRKAEELNRVMGCVGKILGALLTIVSVVAAAFSGGASL
                     ALAAVGLALMVTDAIVQAATGNSFMEQALNPIMKAVIEPLIKLLSDAFTKMLEGLGVD
                     SKKAKMIGSILGAIAGALVLVAAVVLVATVGKQAAAKLAENIGKIIGKTLTDLIPKFL
                     KNFSSQLDDLITNAVARLNKFLGAAGDEVISKQIISTHLNQAVLLGESVNSATQAGGS
                     VASAVFQNSASTNLADLTLSKYQVEQLSKYISEAIEKFGQLQEVIADLLASMSNSQAN
                     RTDVAKAILQQTTA"
RBS                  4224. .4226
                     /note="put. ribosome binding site"
CDS                  4231. .5379
                     /note="ipaC gene (AA 1-382)"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA33382.1"
                     /db_xref="PID:g47056"
                     /db_xref="GI:47056"
                     /db_xref="SWISS-PROT:P18012"
                     /translation="MLQKQFCNKLLLDTNKENVMEIQNTKPTQILYTDISTKQTQSSS
                     ETQKSQNYQQIAAHIPLNVGKNPVLTTTLNDDQLLKLSEQVQHDSEIIARLTDKKMKD
                     LSEMSHTLTPENTLDISSLSSNAVSLIISVAVLLSALRTAETKLGSQLSLIAFDATKS
                     AAENIVRQGLAALSSSITGAVTQVGITGIGAKKTHSGISDQKGALRKNLATAQSLEKE
                     LAGSKLGLNKQIDTNITSPQTNSSTKFLGKNKLAPDNISLSTEHKTSLSSPDISLQDK
                     IDTQRRTYELNTLSAQQKQNIGRATMETSAVAGNISTSGGRYASALEEEEQLISQASS
                     KQAEEASQVSKEASQATNQLIQKLLNIIDSINQSKNSTASQIAGNIRA"
promoter             5376. .5381
                     /note="-35 region for 4.5kb transcript"
promoter             5396. .5401
                     /note="-10 region for 4.5kb transcript"
RBS                  5421. .5423
                     /note="put. ribosome binding site"
CDS                  5430. .6428
                     /note="ipaD gene (AA 1-332)"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA33383.1"
                     /db_xref="PID:g47057"
                     /db_xref="GI:47057"
                     /db_xref="SWISS-PROT:P18013"
                     /translation="MNITTLTNSISTSSFSPNNTNGSSTETVNSDIKTTTSSHPVSSL
                     TMLNDTLHNIRTTNQALKKELSQKTLTKTSLEEIALHSSQISMDVNKSAQLLDILSRH
                     EYPINKDARELLHSAPKEAELDGDQMISHRELWAKIANSINDINEQYLKVYEHAVSSY
                     TQMYQDFSAVLSSLAGWISPGGNDGNSVKLQVNSLKKALEELKEKYKDKPLYPANNTV
                     SQEQANKWLTELGGTIGKVSQKNGGYVVSINMTPIDNMLKSLDNLGGNGEVVLDNAKY
                     QAWNAGFSAEDETMKNNLQTLVQKYSNANSIFDNLVKVLSSTISSCTDTDKLFLHF"
BASE COUNT      2845 a    1501 c    1246 g    2071 t
ORIGIN


alignment_scores:
           Quality:  379.00              Length:    404
             Ratio:  1.452                 Gaps:      6
 Percent Similarity:  64.604   Percent Identity: 29.455

alignment_block:
 US-09-068-804-6 x SFPVR2   ..

 Align seg 1/1  to: SFPVR2  from: 1  to: 7663

       9 AsnProAlaAlaTyrLeuAsnAsnHisSerValGluAsnSerSerGlnTh 25
         :::|||:::      |||   ::::::         :::   :::|||::
    4306 AAACCAACCCAGATTTTATATACAGATATATCCACAAAACAAACTCAAAG 4355

      25 rAlaSerGlnSerValSerAlaLysAsp.....IleLeuAsnSerI 39
         ::::|||:::::      :::::::::         |||   :::   |
    4356 TTCTTCCGAAACACAAAAATCACAAAATTATCAGCAGATTGCAGCGCATA 4405

      39 leGlyIleSerSerSerLysValSerAspLeuGlyLeuSerProThrLeu 55
         ||   ::::::   :::|||      :::      |||:::::||||||
    4406 TTCCACTTAATGTCGGTAAA......AATCCCGTATTAACAACCACATTA 4449

      56 SerAlaProAlaProGlyValLeuThrGlnThrProGlyThrIleThrSe 72
         ::::::
    4450 AATGAT.....................................GATCA 4460

      72 rPheLeuLysAlaSerIleGlnAsnThrAspMetAsnGlnAspLeuAsnA 89
          :::||||||   |||   |||      :::   ::::::   ::::::
    4461 ACTTTTAAAGTTATCAGAGCAGGTTCAGCATGATTCAGAAATCATTGCTC 4510

      89 laLeuAlaAsnAsnValThrThrLysAlaAsnGluValValGlnThrGln 105
           |||:::::::::      :::      :::|||:::   :::|||
    4511 GCCTTACTGACAAAAAGATGAAAGATCTTTCAGAGATGAGTCACACCCTT 4560

     106 LeuArgGluGlnGlnAlaGluValGlyLysPhePheAspIleSerGlyMe 122
               |||::::                 :::||||||||||:::::
    4561 ACTCCAGAGAACACT.................CTGGATATTTCCAGTCT 4592

     122 tSerSerSerAlaValAlaLeuLeuAlaAlaAlaAsnThrLeuMetLeuI 139
         :||||||:::|||||:::|||:::   :::::::::::|||:::   :
    4593 TTCTTCTAATGCTGTTTCTTTAATTATTAGTGTAGCCGTTCTACTTTCTG 4642

     139 hrLeuAsnGlnAlaAspSerLysLeuSerGlyLysLeuSerLeuValSer 155
         ::|||      |||:::::|||||:::::::::|||||||||:::::
    4643 CTCTCCGCACTGCAGAAACTAAATTGGGCTCTCAATTGTCATTGATTGCG 4692

     156 PheAspAlaAlaLysThrThrAlaSerSerMetMetArgGluGlyMetAs 172
         |||||||||:::|||:::::|||   :::::::::|||:::|||:::::
    4693 TTCGATGCTACAAAATCAGCTGCAGAGAACATTGTTCGGCAAGGCCTGGC 4742

     172 nAlaLeuSerGlySerIleSerGlnSerAlaLeuGlnLeuGlyIleThrG 189
         :|||||||||:::|||||:::   ::::::   |||:::|||||||||
    4743 AGCCCTATCATCAAGCATTACTGGAGCAGTCACACAAGTAGGTATAACGG 4792

     189 lyValGlyAlaLysLeuGluTyrLysGlyLeuGlnAsnGluArgGlyAla 205
         ||:::|||||||||      :::   |||:::   :::::::::||||||
    4793 GTATCGGTGCCAAAAAAACGCATTCAGGGATTAGCGACCAAAAAGGAGCC 4842

     206 LeuLysHisAsnAlaAlaLysIleAspLysLeuThrThrGluSerHisSe 222
         |||:::   |||   |||:::   :::   |||   :::|||      ::
    4843 TTAAGAAAGAACCTTGCCACTGCTCAATCTCTTGAAAAAGAGCTTGCAGG 4892

     222 rIleLysAsnValLeuAsnGlyGlnAsnSerValLysLeuGlyAlaGluG 239
         :   |||      ||||||   |||   ::::::::::::   :::
```

```
alignment_scores:
            Quality:  380.50              Length:    355
              Ratio:  1.605                 Gaps:      4
 Percent Similarity:  66.761   Percent Identity: 32.394

alignment_block:
 US-09-068-804-6 x SHFIPABCD   ..

 Align seg 1/1  to: SHFIPABCD  from: 1  to: 5255

      55 LeuSerAlaProAlaProGlyValLeuThrGlnThrProGlyThrIleTh 71
         |||::::::              |||||||||   |||   :::
    2964 CTTAATGTCGGTAAAAATCCCGTATTAACAACCACATTAAATGAT...GA 3010

      71 rSerPheLeuLysAlaSerIleGlnAsnThrAspMetAsnGlnAspLeuA 88
           :::|||||||   |||   |||      :::   ::::::   ::::
    3011 TCAACTTTTAAAGTTATCAGAGCAGGTTCAGCATGATTCAGAAATCATTG 3060

      88 snAlaLeuAlaAsnAsnValThrThrLysAlaAsnGluValValGlnThr 104
         ::   |||:::::::::      :::      :::|||:::   :::|||
    3061 CTCGCCTTACTGACAAAAAGATGAAAGATCTTTCAGAGATGAGTCACACC 3110

     105 GlnLeuArgGluGlnGlnAlaGluValGlyLysPhePheAspIleSerGl 121
                |||:::                        :::|||||||||::
    3111 CTTACTCCAGAGAACACT.................CTGGATATTTCCAG 3142

     121 yMetSerSerSerAlaValAlaLeuLeuAlaAlaAlaAsnThrLeuMetL 138
         ::::|||||||:::||||||:::|||:::   :::::::::::::|||:::
    3143 TCTTTCTTCTAATGCTGTTTCTTTAATTATTAGTGTAGCCGTTCTACTTT 3192

     138 euThrLeuAsnGlnAlaAspSerLysLeuSerGlyLysLeuSerLeuVal 154
           :::|||      |||::::::||||||:::::::::||||||||:::
    3193 CTGCTCTCCGCACTGCAGAAACTAAATTGGGCTCTCAATTGTCATTGATT 3242

     155 SerPheAspAlaAlaLysThrThrAlaSerSerMetMetArgGluGlyMe 171
         :::|||||||||:::|||:::::|||   :::::::::|||:::|||::
    3243 GCGTTCGATGCTACAAAATCAGCTGCAGAGAACATTGTTCGGCAAGGCCT 3292

     171 tAsnAlaLeuSerGlySerIleSerGlnSerAlaLeuGlnLeuGlyIleT 188
         ::::|||||||||:::||||||:::   ::::::   |||:::|||||||
    3293 GGCAGCCCTATCATCAAGCATTACTGGAGCAGTCACACAAGTAGGTATAA 3342

     188 hrGlyValGlyAlaLysLeuGluTyrLysGlyLeuGlnAsnGluArgGly 204
         |||||:::|||||||||      :::   |||:::   ::::::::|||
    3343 CGGGTATCGGTGCCAAAAAAACGCATTCAGGGATTAGCGACCAAAAAGGA 3392

     205 AlaLeuLysHisAsnAlaAlaLysIleAspLysLeuThrThrGluSerHi 221
         ||||||:::   |||   |||:::   :::   |||   :::|||
    3393 GCCTTAAGAAAGAACCTTGCCACTGCTCAATCTCTTGAAAAAGAGCTTGC 3442

     221 sSerIleLysAsnValLeuAsnGlyGlnAsnSerValLysLeuGlyAlaG 238
          :::   |||      ||||||   |||   ::::::::::::   :::
    3443 AGGTTCTAAATTAGGGTTAAATAAACAAATAGATACAAATATCACCTCAC 3492

     238 luGlyValAspSerLeuLysSerLeuAsnMetLysLysThrGlyThrAsp 254
           ::::::|||
    3493 CACAAACTAACTCT..................................... 3506

     255 AlaThrLysAsnLeuAsnAspAlaThrLeuLysSerAsnAlaGlyThrSe 271
         :::||||||   |||:::   :::::|||   ::::::
    3507 AGCACAAAATTTTTAGGTAAAAATAAACTGGCGCCAGAT.......... 3545

     271 rAlaThrGluSerLeuGlyIleLysAspSerAsnLysGlnIleSerProG 288
                                      |||      :::|||:::|
    3546 ..............................AATATATCCCTGTCAACTG 3564

     288 luHisGlnAlaIleLeuSerLysArgLeuGluSerValGluSerAspIle 304
         |||||::::::   ||||||           |||:::::::::   |||
    3565 AACATAAAACTTCTCTTAGTTCTCCCGATATTTCTTTGCAGGATAAAATT 3614

     305 ArgLeuGluGlnAsnThrMetAspMetThrArgIleAspAlaArgLysMe 321
                 ::::::   |||   :::::::::   ::::::|||::::::::
    3615 GACACCCAGAGAAGAACTTACGAGCTCAATACCCTTTCTGCGCAGCAAAA 3664

     321 tGlnMetThrGlyAspLeuIleMetLysAsnSerValThrValGlyGlyI 338
         :|||   :::|||      :::|||   :::|||:::::::::|||:::|
    3665 ACAAAACATTGGCCGTGCAACAATGGAAACATCAGCCGTTGCTGGTAATA 3714

     338 leAlaGlyAlaSerGlyGlnTyrAlaAlaThrGlnGluArgSerGluGln 354
         ||:::   ::::::|||:::||||||::::::   |||      ||||||
    3715 TATCCACATCAGGAGGGCGTTATGCATCTGCTCTTGAAGAAGAAGAACAA 3764

     355 GlnIleSerGlnValAsnAsnArgValAlaSerThrAlaSerAspGluAl 371
           ||||||||:::::::::::   |||      ||||||:::   ::
    3765 CTAATCAGTCAGGCCAGCAGTAAACAAGCAGAGGAAGCATCCCAAGTATC 3814

     371 aArgGluSerSerArgLysSerThrSerLeuIleGlnGluMetLeuLysT 388
         :::|||:::|||:::   ::::::   |||||||||   :::|||::::
    3815 TAAAGAAGCATCCCAAGCGACAAATCAATTAATACAAAAATTATTGAATA 3864

     388 hrMetGluSerIleAsnGlnSerLysAlaSerAlaLeuAlaAlaIleAla 404
         ::::::::|||||||||||||||||||:::||||||   :::   ||||||
    3865 TAATTGACAGCATCAACCAATCAAAGAATTCGGCAGCCAGTCAGATTGCT 3914

     405 GlyAsnIleArgAla 409
         |||||||||||||||
    3915 GGTAACATTCGAGCT 3929

seq_name: gb_ba1:SFPVR2

seq_documentation_block:
LOCUS       SFPVR2          7663 bp    DNA             BCT       12-SEP-1993
DEFINITION  Shigella flexineri plasmid YSH6000 virulence region-2 DNA
            containing 24k gene, 18k gene, ipaB gene, ipaC gene and ipaD gene.
ACCESSION   X15319
NID         g47052
VERSION     X15319.1  GI:47052
KEYWORDS    ipaB gene; ipaC gene; ipaD gene; virulence proteins.
SOURCE      Enterobacteriaceae.
  ORGANISM  Enterobacteriaceae
            Bacteria; Proteobacteria; gamma subdivision.
REFERENCE   1  (bases 1 to 7663)
  AUTHORS   Sasakawa,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAY-1989) Sasakawa C., Institute of Medical Science,
            University of Tokyo, Shirokanedai 4-6-1, Minato-ku Tokyo 108, Japan
REFERENCE   2  (bases 1010 to 6538)
  AUTHORS   Sasakawa,C., Adler,B., Tobe,T., Okada,N., Nagai,S., Komatsu,K. and
            Yoshikawa,M.
  TITLE     Functional organization and nucleotide sequence of virulence
            Region-2 on the large virulence plasmid in Shigella flexneri 2a
  JOURNAL   Mol. Microbiol. 3 (9), 1191-1201 (1989)
  MEDLINE   90014179
COMMENT     Region 2 has three promoters that transcribe RNAs of 4.0, 4.5 and
            7.5kb. The 4.0 kb transcript produces the 24,18kD, IpaB and IpaC
            proteins. The 4.5 kb transcript produces the ipaD protein. The
            7.5kb transcript covers the entire region 2.
            Data kindly reviewed (20-NOV-1989) by Sasakawa C.
FEATURES             Location/Qualifiers
     source          1. .7663
                     /organism="Enterobacteriaceae"
                     /plasmid="YSH6000"
                     /db_xref="taxon:543"
     promoter        1291. .1296
                     /note="-35 region for 4kb transcript"
     promoter        1312. .1318
                     /note="-10 region for 4kb transcript"
     RBS             1360. .1362
                     /note="put. ribosome binding site"
     CDS             1370. .1996
                     /note="24k gene (AA 1-208)"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA33379.1"
```

```
      |||||:::|||||||||      :::   |||:::   :::::::::|||
   3343 CGGGTATCGGTGCCAAAAAAACGCATTCAGGGATTAGCGACCAAAAAGGA 3392

    205 AlaLeuLysHisAsnAlaAlaLysIleAspLysLeuThrThrGluSerHi 221
        ||||||:::   |||   |||:::   :::   |||   :::|||
   3393 GCCTTAAGAAAGAACCTTGCCACTGCTCAATCTCTTGAAAAAGAGCTTGC 3442

    221 sSerIleLysAsnValLeuAsnGlyGlnAsnSerValLysLeuGlyAlaG 238
         :::   |||     ||||||   |||   ::::::::::::   :::
   3443 AGGTTCTAAATTAGGGTTAAATAAACAAATAGATACAAATATCACCTCAC 3492

    238 luGlyValAspSerLeuLysSerLeuAsnMetLysLysThrGlyThrAsp 254
           :::::::|||
   3493 CACAAACTAACTCT..................................... 3506

    255 AlaThrLysAsnLeuAsnAspAlaThrLeuLysSerAsnAlaGlyThrSe 271
        :::||||||   |||:::   :::::::|||   ::::::
   3507 AGCACAAAATTTTTAGGTAAAAATAAACTGGCGCCAGAT.......... 3545

    271 rAlaThrGluSerLeuGlyIleLysAspSerAsnLysGlnIleSerProG 288
                                       |||      :::|||:::|
   3546 ..............................AATATATCCCTGTCAACTG 3564

    288 luHisGlnAlaIleLeuSerLysArgLeuGluSerValGluSerAspIle 304
        |||||::::::   ||||||            |||:::::::::   |||
   3565 AACATAAAACTTCTCTTAGTTCTCCCGATATTTCTTTGCAGGATAAAATT 3614

    305 ArgLeuGluGlnAsnThrMetAspMetThrArgIleAspAlaArgLysMe 321
           ::::::   |||   :::::::::   ::::::|||:::::::::
   3615 GACACCCAGAGAAGAACTTACGAGCTCAATACCCTTTCTGCGCAGCAAAA 3664

    321 tGlnMetThrGlyAspLeuIleMetLysAsnSerValThrValGlyGlyI 338
        :|||   :::|||      :::|||   :::|||::::::::|||:::|
   3665 ACAAAACATTGGCCGTGCAACAATGGAAACATCAGCCGTTGCTGGTAATA 3714

    338 leAlaGlyAlaSerGlyGlnTyrAlaAlaThrGlnGluArgSerGluGln 354
        ||:::   ::::::|||:::||||||::::::   |||      ||||||
   3715 TATCCACATCAGGAGGGCGTTATGCATCTGCTCTTGAAGAAGAAGAACAA 3764

    355 GlnIleSerGlnValAsnAsnArgValAlaSerThrAlaSerAspGluAl 371
           ||||||||:::::::::::::   |||      ||||||:::   ::
   3765 CTAATCAGTCAGGCCAGCAGTAAACAAGCAGAGGAAGCATCCCAAGTATC 3814

    371 aArgGluSerSerArgLysSerThrSerLeuIleGlnGluMetLeuLysT 388
        ::::|||:::|||:::   ::::::   |||||||||   :::|||:::::
   3815 TAAAGAAGCATCCCAAGCGACAAATCAATTAATACAAAAATTATTGAATA 3864

    388 hrMetGluSerIleAsnGlnSerLysAlaSerAlaLeuAlaAlaIleAla 404
        ::::::::|||||||||||||||||||:::|||||   :::   ||||||
   3865 TAATTGACAGCATCAACCAATCAAAGAATTCGGCAGCCAGTCAGATTGCT 3914

    405 GlyAsnIleArgAla 409
        |||||||||||||||
   3915 GGTAACATTCGAGCT 3929

seq_name: gb_ba1:SHFIPABCD

seq_documentation_block:
LOCUS       SHFIPABCD    5255 bp    DNA             BCT       19-MAR-1991
DEFINITION  S.flexneri ipaBCD region encoding invasion plasmid antigens.
ACCESSION   J04117
NID         g152740
VERSION     J04117.1  GI:152740
KEYWORDS    ipaA gene; ipaB gene; ipaC gene; ipaD; membrane protein.
SOURCE      S.flexeri (strain M90T-W) DNA, clone pHC17.
  ORGANISM  Shigella flexneri
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Shigella.
REFERENCE   1  (bases 1 to 5255)
  AUTHORS   Venkatesan,M.M., Buysse,J.M. and Kopecko,D.J.
  TITLE     Characterization of invasion plasmid antigen genes (ipaBCD) from
            Shigella flexneri
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 85, 9317-9321 (1988)
  MEDLINE   89057927
COMMENT     Draft entry and clean copy of sequence [1] kindly submitted by
            M.M.Venkatesan, 03-JAN-1989.
FEATURES             Location/Qualifiers
     source          1. .5255
                     /organism="Shigella flexneri"
                     /db_xref="taxon:623"
     CDS             606. .1073
                     /note="18 kD protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA26521.1"
                     /db_xref="PID:g152741"
                     /db_xref="GI:152741"
                     /translation="MSLNITENESISTAVIDAINSGATLKDINAIPDDMMDDIYSYAY
                     DFYNKGRIEEAEVFFRFLCIYDFYNVDYIMGLAAIYQIKEQFQQAADLYAVAFALGKN
                     DYTPVFHTGQCQLRLKAPLKAKECFELVIQHSNDEKLKIKAQSYLDAIQDIKE"
     CDS             1079. .2821
                     /note="IpaB protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA26522.1"
                     /db_xref="PID:g152742"
                     /db_xref="GI:152742"
                     /translation="MHNVSTTTTGFPLAKILTSTELGDNTIQAANDAANKLFSLTIAD
                     LTANQNINTTNAHSTSNILIPELKAPKSLNASSQLTLLIGNLIQILGEKSLTALTNKI
                     TAWKSQQQARQQKNLEFSDKINTLLSETEGLTRDYEKQINKLKNADSKIKDLENKINQ
                     IQTRLSNLDPESPEKKKLSREEIQLTIKKDAAVKDRTLIEQKTLSIHSKLTDKSMQLE
                     KEIDSFSAFSNTASAEQLSTQQKSLTGLASVTQLMATFIQLVGKNNEESLKNDLALFQ
                     SLQESRKTEMERKSDEYAAEVRKAEELNRVMGCVGKILGALLTIVSVVAAAFSGGASL
                     ALAAVGLALMVTDAIVQAATGNSFMEQALNPIMKAVIEPLIKLLSDAFTKMLEGLGVD
                     SKKAKMIGSILGAIAGALVLVAAVVLVATVGKQAAAKLAENIGKIIGKTLTDLIPKFL
                     KNFSSQLDDLITNAVARLNKFLGAAGDEVISKQIISTHLNQAVLLGESVNSATQAGGS
                     VASAVFQNSASTNLADLTLSKYQVEQLSKYISEAIEKFGQLQEVIADLLASMSNSQAN
                     RTDVAKAILQQTTA"
     CDS             2784. .3932
                     /note="IpaC protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA26523.1"
                     /db_xref="PID:g152743"
                     /db_xref="GI:152743"
                     /translation="MLQKQFCNKLLLDTNKENVMEIQNTKPTQTLYTDISTKQTQSSS
                     ETQKSQNYQQIAAHIPLNVGKNPVLTTTLNDDQLLKLSEQVQHDSEIIARLTDKKMKD
                     LSEMSHTLTPENTLDISSLSSNAVSLIISVAVLLSALRTAETKLGSQLSLIAFDATKS
                     AAENIVRQGLAALSSSITGAVTQVGITGIGAKKTHSGISDQKGALRKNLATAQSLEKE
                     LAGSKLGLNKQIDTNITSPQTNSSTKFLGKNKLAPDNISLSTEHKTSLSSPDISLQDK
                     IDTQRRTYELNTLSAQQKQNIGRATMETSAVAGNISTSGGRYASALEEEEQLISQASS
                     KQAEEASQVSKEASQATNQLIQKLLNIIDSINQSKNSAASQIAGNIRA"
     CDS             3983. .4981
                     /note="IpaD protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA26524.1"
                     /db_xref="PID:g152744"
                     /db_xref="GI:152744"
                     /translation="MNITTLTNSISTSSFSPNNINGSSTETVNSDIKTTTSSHPVSSL
                     TMLNDTLHNIRTTNQALKKELSQKTLTKTSLEEIALHSSQISMDVNKSAQLLDILSRN
                     EYPINKDARELLHSAPKEAELDGDQMISHRELWAKIANSINDINEQYLKVYEHAVSSY
                     TQMYQDFSAVLSSLAGWISPGGNDGNSVKLQVNSLKKALEELKEKYKDKPLYPANNTV
                     SQEQANKWLTELGGTIGKVSQKNGGYVVSINMTPIDNMLKSLDNLGGNGEVVLDNAKY
                     QAWNAGFSAEDETMKNNLQTLVQKYSNANSIFDNLVKVLSSTISSCTDTDKLFLHF"
     CDS             4990. .>5255
                     /note="IpaA protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA26525.1"
                     /db_xref="PID:g551981"
                     /db_xref="GI:551981"
                     /translation="MHNVNNTQAPTFLYKATSPSSTEYSELKSKISDIHSSQTSLKTP
                     ASVSEKENFATSFNQKCLDFLFSSSGKEDVLRSIYSNSMNAYAK"
BASE COUNT     1950 a    1037 c     889 g    1379 t
ORIGIN
```

```
           Baltimore, MD 21201.
FEATURES             Location/Qualifiers
     source          1. .4761
                     /organism="Plasmid pWR100"
                     /plasmid="Plasmid pWR100"
                     /specific_host="M90T"
                     /db_xref="taxon:2661"
                     /clone="HB101/pHS[4108,5357]."
     CDS             <1. .549
                     /note="ORF"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA98422.1"
                     /db_xref="PID:g155328"
                     /db_xref="GI:155328"
                     /translation="KLVKKILSVFNLKQRFPQKNFGCPVNINKIRDSVIDKIKDSNSG
                     NQLFCWMSQERTTYVSSMINRSIDEMAIHNGVVLTSDNKKNIFAAIEKKFPDIKLDEK
                     SAQTSISHTALNEIASSGLRAKILKRYSSDMDLFNTQMKDLTNLVSSSVYDKIFNEST
                     KVLQIEISAEVLKAVYRQSNTN"
     -10_signal      556. .561
                     /gene="ippI"
     gene            556. .1073
                     /gene="ippI"
     RBS             595. .600
                     /gene="ippI"
     CDS             606. .1073
                     /gene="ippI"
                     /codon_start=1
                     /transl_table=11
                     /product="antigen"
                     /protein_id="AAA98423.1"
                     /db_xref="PID:g155329"
                     /db_xref="GI:155329"
                     /translation="MSLNITENESISTAVIDAINSGATLKDINAIPDDMMDDIYSYAY
                     DFYNKGRIEEAEVFFRFLCIYDFYNVDYIMGLAAIYQIKEQFQQAADLYAVAFALGKN
                     DYTPVFHTGQCQLRLKAPLKAKECFELVIQHSNDEKLKIKAQSYLDAIQDIKE"
     gene            1065. .2821
                     /gene="ipaB"
     RBS             1065. .1070
                     /gene="ipaB"
     CDS             1079. .2821
                     /gene="ipaB"
                     /codon_start=1
                     /transl_table=11
                     /product="antigen"
                     /protein_id="AAA98424.1"
                     /db_xref="PID:g155330"
                     /db_xref="GI:155330"
                     /translation="MHNVSTTTGFPLAKILTSTELGDNTIQAANDAANKLFSLTIAD
                     LTANQNINTTNAHSTSNILIPELKAPKSLNASSQLTLLIGNLIQILGEKSLTALTNKI
                     TAWKSQQQARQQKNLEFSDKINTLLSETEGLTRDYEKQINKLKNADSKIKDLENKINQ
                     IQTRLSELDPESPEKKKLSREEIQLTIKKDAAVKDRTLIEQKTLSIHSKLTDKSMQLE
                     KEIDSFSAFSNTASAEQLSTQQKSLTGLASVTQLMATFIQLVGKNNEESLKNDLALFQ
                     SLQESRKTEMERKSDEYAAEVRKAEELNRVMGCVGKILGALLTIVSVVAAAFSGGASL
                     ALAAVGLALMVTDAIVQAATGNSFMEQALNPIMKAVIEPLIKLLSDAFTKMLEGLGVD
                     SKKAKMIGSILGAIAGALVLVAAVVLVATVGKQAAAKLAENIGKIIGKTLTDLIPKFL
                     KNFSSQLDDLITNAVARLNKFLGAAGDEVISKQIISTHLNQAVLLGESVNSATQAGGS
                     VASAVFQNSASTNLADLTLSKYQVEQLSKYISEAIEKFGQLQEVIADLLASMSNSQAN
                     RTDVAKAILQQTTA"
     gene            2723. .3932
                     /gene="ipaC"
     -35_signal      2723. .2728
                     /gene="ipaC"
     -10_signal      2745. .2750
                     /gene="ipaC"
     RBS             2829. .2834
                     /gene="ipaC"
     CDS             2841. .3932
                     /gene="ipaC"
                     /codon_start=1
                     /transl_table=11
                     /product="antigen"
                     /protein_id="AAA98425.1"
                     /db_xref="PID:g155331"
                     /db_xref="GI:155331"
                     /translation="MEIQNTKPTQTLYTDISTKQTQSSSETQKSQNYQQIAAHIPLNV
                     GKNPVLTTTLNDDQLLKLSEQVQHDSEIIARLTDKKMKDLSEMSHTLTPENTLDISSL
                     SSNAVSLIISVAVLLSALRTAETKLGSQLSLIAFDATKSAAENIVRQGLAALSSSITG
                     AVTQVGITGIGAKKTHSGISDQKGALRKNLATAQSLEKELAGSKLGLNKQIDTNITSP
                     QTNSSTKFLGKNKLAPDNISLSTEHKTSLSSPDISLQDKIDTQRRTYELNTLSAQQKQ
                     NIGRATMETSAVAGNISTSGGRYASALEEEEQLISQASSKQAEEASQVSKEASQATNQ
                     LIQKLLNIIDSINQSKNSAASQIAGNIRA"
     RBS             3971. .3977
                     /gene="ipaD"
     gene            3971. .4761
                     /gene="ipaD"
     CDS             3983. .>4761
                     /gene="ipaD"
                     /codon_start=1
                     /transl_table=11
                     /product="antigen"
                     /protein_id="AAA98426.1"
                     /db_xref="PID:g552045"
                     /db_xref="GI:552045"
                     /translation="MNITTLTNSISTSSFSPNNTNGSSTETVNSDIKTTTSSHPVSSL
                     TMLNDTLHNIRTTNQALKKELSQKTLTKTSLEEIALHSSQISMDVNKSAQLLDILSRN
                     EYPINKDARELLHSAPKEAELDGDQMISHRELWAKIANSINDINEQYLKVYEHAVSSY
                     TQMYQDFSAVLSSLAGWISPGGNDGNSVKLQVNSLKKALEELKEKYKDKPLYPANNTV
                     SQEQANKWLTELGGTIGKVSQKNGGYVVSINMTPIDNMLKS"
BASE COUNT     1769 a    952 c    805 g   1235 t
ORIGIN


alignment_scores:
            Quality:  380.50            Length:    355
              Ratio:  1.605               Gaps:      4
 Percent Similarity:  66.761   Percent Identity: 32.394

alignment_block:
 US-09-068-804-6 x W10PPIBCA   ..

 Align seg 1/1  to: W10PPIBCA  from: 1  to: 4761

      55 LeuSerAlaProAlaProGlyValLeuThrGlnThrProGlyThrIleTh 71
         |||:::::::            |||||||||   |||   :::
    2964 CTTAATGTCGGTAAAAATCCCGTATTAACAACCACATTAAATGAT...GA 3010

      71 rSerPheLeuLysAlaSerIleGlnAsnThrAspMetAsnGlnAspLeuA 88
         :::||||||   |||   |||      :::   ::::::   ::::
    3011 TCAACTTTTAAAGTTATCAGAGCAGGTTCAGCATGATTCAGAAATCATTG 3060

      88 snAlaLeuAlaAsnAsnValThrThrLysAlaAsnGluValValGlnThr 104
         ::   |||:::::::::      :::      :::|||:::   :::|||
    3061 CTCGCCTTACTGACAAAAAGATGAAAGATCTTTCAGAGATGAGTCACACC 3110

     105 GlnLeuArgGluGlnGlnAlaGluValGlyLysPhePheAspIleSerGl 121
                |||:::                    :::||||||||||::
    3111 CTTACTCCAGAGAACACT.................CTGGATATTTCCAG 3142

     121 yMetSerSerSerAlaValAlaLeuLeuAlaAlaAlaAsnThrLeuMetL 138
         ::::||||||:::||||||:::|||:::   ::::::::::::|||:::
    3143 TCTTTCTTCTAATGCTGTTTCTTTAATTATTAGTGTAGCCGTTCTACTTT 3192

     138 euThrLeuAsnGlnAlaAspSerLysLeuSerGlyLysLeuSerLeuVal 154
           :::|||      |||:::::||||||:::::::::|||||||||:::
    3193 CTGCTCTCCGCACTGCAGAAACTAAATTGGGCTCTCAATTGTCATTGATT 3242

     155 SerPheAspAlaAlaLysThrThrAlaSerSerMetMetArgGluGlyMe 171
         :::|||||||||:::|||:::::|||   :::::::::|||:::|||::
    3243 GCGTTCGATGCTACAAAATCAGCTGCAGAGAACATTGTTCGGCAAGGCCT 3292

     171 tAsnAlaLeuSerGlySerIleSerGlnSerAlaLeuGlnLeuGlyIleT 188
         ::::|||||||||:::|||||:::   ::::::   |||:::|||||||
    3293 GGCAGCCCTATCATCAAGCATTACTGGAGCAGTCACACAAGTAGGTATAA 3342

     188 hrGlyValGlyAlaLysLeuGluTyrLysGlyLeuGlnAsnGluArgGly 204
```

```
NID         g1513311
VERSION     U66877.1  GI:1513311
KEYWORDS    .
SOURCE      Salmonella dublin.
  ORGANISM  Salmonella dublin
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Salmonella.
REFERENCE   1  (bases 1 to 2995)
  AUTHORS   Mullan,P.B., Gautier,A.V., Wood,M.W., Edwards,M.H., Jones,B.V. and
            Galyov,E.E.
  TITLE     Sip invasins of Salmonella dublin are involved in macrophage
            cytotoxicity and mouse virulence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2995)
  AUTHORS   Mullan,P.B., Gautier,A.V., Wood,M.W., Edwards,M.H., Jones,B.V. and
            Galyov,E.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-AUG-1996) Institute for Animal Health, Compton,
            Newbury, Berkshire, RG20 7NN, UK
FEATURES             Location/Qualifiers
     source          1. .2995
                     /organism="Salmonella dublin"
                     /strain="2229"
                     /db_xref="taxon:593"
     gene            131. .628
                     /gene="sicA"
     CDS             131. .628
                     /gene="sicA"
                     /codon_start=1
                     /transl_table=11
                     /product="SicA"
                     /protein_id="AAB06794.1"
                     /db_xref="PID:g1513312"
                     /db_xref="GI:1513312"
                     /translation="MDYQNNVSEERVAEMIWDAVSEGATLKDVHGIPQDMMDGLYAHA
                     YEFYNQGRLDEAETFFRFLCIYDFYNPDYTMGLAAVCQLKKQFQKACDLYAVAFTLLK
                     NDYRPVFFTGQCQLLMRKAAKARQCFELVNERTEDESLRAKALVYLEALKTAETEQHS
                     EQEKE"
     gene            631. .2412
                     /gene="sipB"
     CDS             631. .2412
                     /gene="sipB"
                     /function="virulence-associated secreted protein"
                     /codon_start=1
                     /transl_table=11
                     /product="SipB"
                     /protein_id="AAB06795.1"
                     /db_xref="PID:g1513313"
                     /db_xref="GI:1513313"
                     /translation="MVNDASSISRSGYTQNPRLAEAAFEGVRKNTDFLKAADKAFKDV
                     VATKAGDLKAGTKSGESAINTVGLKPPTDAAREKLSSEGQLTLLLGKLMTLLGDVSLS
                     QLESRLAVWQAMIESQKEMGIQVSKEFQTALGEAQEATDLYEASIKKTDTAKSVYDAA
                     TKKLTQAQNKLQSLDPADPGYAQAEAAVEQAGKEATEAKEALDKATDATVKAGTDAKA
                     KAEKADNILTKFQGTANAASQNQVSQGEQDNLSNVARLTMLMAMFIEIVGKNTEESLQ
                     NDLALFNALQEGRQAEMEKKSAEFQEETRKAEETNRIMGCIGKVLGALLTIVSVVAAV
                     FTGGASLALAAVGLAVMVADEIVKAATGVSFIQQALNPIMEHVLKPLMELIGKAITKA
                     LEGLGVDKKTAEMAGSIVGAIVAAIAMVAVIVVVAVVGKGAAAKLGNALSKMMGETIK
                     KLVPNVLKQLAQNGSKLFTQGMQRITSGLGNVGSKMGLQTNALSKELVGNTLNKVALG
                     MEVTNTAAQSAGGVAEGVFIKNASEALADFMLARFAMDQIQQWLKQSVEIFGENQKVT
                     AELQKAMSSAVQQNADASRFILRQSRA"
BASE COUNT      851 a    614 c    847 g    683 t
ORIGIN


alignment_scores:
            Quality:  739.00            Length:    185
              Ratio:  3.995               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-068-804-6 x SDU66877   ..

 Align seg 1/1  to: SDU66877  from: 1  to: 2995
```

```
    1 MetLeuIleSerAsnValGlyIleAsnProAlaAlaTyrLeuAsnAsnHi 17
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 2440 ATGTTAATTAGTAATGTGGGAATAAATCCCGCCGCTTATTTAAATAATCA 2489

   17 sSerValGluAsnSerSerGlnThrAlaSerGlnSerValSerAlaLysA 34
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 2490 TTCAGTTGAGAATAGTTCACAGACAGCTTCGCAATCCGTTAGCGCTAAAG 2539

   34 spIleLeuAsnSerIleGlyIleSerSerSerLysValSerAspLeuGly 50
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 2540 ATATTCTGAATAGTATTGGTATTAGCAGCAGTAAAGTCAGTGACCTGGGG 2589

   51 LeuSerProThrLeuSerAlaProAlaProGlyValLeuThrGlnThrPr 67
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 2590 TTGAGTCCTACACTGAGCGCGCCTGCGCCAGGGGTATTAACGCAAACCCC 2639

   67 oGlyThrIleThrSerPheLeuLysAlaSerIleGlnAsnThrAspMetA 84
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 2640 CGGAACGATCACGTCCTTTTTAAAAGCCAGTATTCAAAATACCGACATGA 2689

   84 snGlnAspLeuAsnAlaLeuAlaAsnAsnValThrThrLysAlaAsnGlu 100
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 2690 ATCAGGATTTGAATGCTCTGGCAAATAATGTCACGACTAAAGCGAATGAG 2739

  101 ValValGlnThrGlnLeuArgGluGlnGlnAlaGluValGlyLysPhePh 117
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 2740 GTTGTGCAAACCCAGTTACGCGAGCAGCAGGCAGAAGTCGGAAAGTTTTT 2789

  117 eAspIleSerGlyMetSerSerSerAlaValAlaLeuLeuAlaAlaAlaA 134
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 2790 TGATATTAGCGGAATGTCTTCCAGTGCCGTTGCGCTGTTGGCTGCCGCGA 2839

  134 snThrLeuMetLeuThrLeuAsnGlnAlaAspSerLysLeuSerGlyLys 150
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 2840 ATACGTTAATGCTGACGTTGAACCAGGCTGATAGCAAACTGTCTGGTAAG 2889

  151 LeuSerLeuValSerPheAspAlaAlaLysThrThrAlaSerSerMetMe 167
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 2890 TTGTCATTAGTCAGCTTTGATGCAGCTAAAACGACAGCAAGCTCCATGAT 2939

  167 tArgGluGlyMetAsnAlaLeuSerGlySerIleSerGlnSerAlaLeuG 184
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 2940 GCGCGAAGGGATGAATGCGTTGTCCGGTAGTATTTCCCAGAGCGCGCTTC 2989

  184 lnLeu 185
      |||||
 2990 AGTTG 2994
```

```
seq_name: gb_ba1:W10PPIBCA

seq_documentation_block:
LOCUS       W10PPIBCA    4761 bp    DNA             BCT       07-MAY-1996
DEFINITION  Plasmid pWR100 (from S.flexneri) antigen ippI, ipaB, ipaC genes,
            complete cds, and ippD gene, 3' end.
ACCESSION   M34849
NID         g155327
VERSION     M34849.1  GI:155327
KEYWORDS    antigen.
SOURCE      Plasmid pWR100 (clone: HB101/pHS[4108,5357].).
  ORGANISM  Plasmid pWR100
            plasmids.
REFERENCE   1  (bases 1 to 4761; 1 to 538; 3987 to 4761)
  AUTHORS   Baudry,B., Kaczorek,M. and Sansonetti,P.J.
  TITLE     Nucleotide sequence of the invasion plasmid antigen B and C genes
            (ipaB and ipaC) of Shigella flexneri
  JOURNAL   Microb. Pathog. 4 (5), 345-357 (1988)
  MEDLINE   89200844
COMMENT     Draft entry and computer-readable sequence for [1] kindly submitted
            by B.Baudry, 05-JUN-1990.
               Center for Vaccine Development
               University of Maryland at Baltimore
               10 South Pine Street
```

```
                     TKKLTQAQNKLQSLDPADPGYAQAEAAVEQAGKEATEAKEALDKATDATVKAGTDAKA
                     KAEKADNILTKFQGTANAASQNQVSQGEQDNLSNVARLTMLMAMFIEIVGKNTEESLQ
                     NDLALFNALQEGRQAEMEKKSAEFQEETRKAEETNRIMGCIGKVLGALLTIVSVVAAV
                     FTGGASLALAAVGLAVMVADEIVKAATGVSFIQQALNPIMEHVLKPLMELIGKAITKA
                     LEGLGVDKKTAEMAGSIVGAIVAAIAMVAVIVVVAVVGKGAAAKLGNALSKMMGETIK
                     KLVPNVLKQLAQNGSKLFTQGMQRITSGLGNVGSKMGLQTNALSKELVGNTLNKVALG
                     MEVTNTAAQSAGGVAEGVFIKNASEALADFMLARFAMDQIQQWLKQSVEIFGENQKVT
                     AELQKAMSSAVQQNADASRFILRQSRA"
     gene            2384. .3613
                     /gene="sipC"
     CDS             2384. .3613
                     /gene="sipC"
                     /codon_start=1
                     /transl_table=11
                     /product="SipC"
                     /protein_id="AAA75170.1"
                     /db_xref="PID:g975296"
                     /db_xref="GI:975296"
                     /translation="MLISNVGINPAAYLNNHSVENSSQTASQSVSAKDILNSIGISSS
                     KVSDLGLSPTLSAPAPGVLTQTPGTITSSLKASIQNTDMNQDLNALANNVTTKANEVV
                     QTQLREQQAEVGKFFDISGMSSSAVALLAAANTLMLTLNQADSKLSGKLSLVSFDAAK
                     TTASSMMREGMNALSGSISQSALQLGITGVGAKLEYKGLQNERGALKHNAAKIDKLTT
                     ESHSIKNVLNGQNSVKLGAEGVDSLKSLNIRKPVPMRRKILMMRRLNLMPEPAPRKVW
                     VLKTVINKVSLNIYILSKRLESVESDIRLEQNYMDITRIDSAQDADDGRSDYEELGHG
                     RWYCRGVRAVRRYSGNVSEQQISQVNNRVASTASDEARESSRKSTSLIQEMLKTMESI
                     NQSKASALAAIAGNIRA"
BASE COUNT     1049 a    759 c   1008 g    806 t
ORIGIN


alignment_scores:
            Quality: 1506.50              Length:    412
              Ratio:  3.748                Gaps:      4
 Percent Similarity:  97.573   Percent Identity: 96.845

alignment_block:
 US-09-068-804-6 x STU25631   ..

 Align seg 1/1  to: STU25631  from: 1  to: 3622

       1 MetLeuIleSerAsnValGlyIleAsnProAlaAlaTyrLeuAsnAsnHi 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2384 ATGTTAATTAGTAATGTGGGAATAAATCCCGCCGCTTATTTAAATAATCA 2433

      17 sSerValGluAsnSerSerGlnThrAlaSerGlnSerValSerAlaLysA 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2434 TTCTGTTGAGAATAGTTCACAGACAGCTTCGCAATCCGTTAGCGCTAAAG 2483

      34 spIleLeuAsnSerIleGlyIleSerSerSerLysValSerAspLeuGly 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2484 ATATTCTGAATAGTATTGGTATTAGCAGCAGTAAAGTCAGTGACCTGGGG 2533

      51 LeuSerProThrLeuSerAlaProAlaProGlyValLeuThrGlnThrPr 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2534 TTGAGTCCTACACTGAGCGCGCCTGCGCCAGGGGTATTAACGCAAACCCC 2583

      67 oGlyThrIleThrSerPheLeuLysAlaSerIleGlnAsnThrAspMetA 84
         ||||||||||||||||   ||||||||||||||||||||||||||||||||
    2584 CGGAACGATCACGTCCTCTTTAAAAGCCAGTATTCAAAATACCGACATGA 2633

      84 snGlnAspLeuAsnAlaLeuAlaAsnAsnValThrThrLysAlaAsnGlu 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2634 ATCAGGATTTGAATGCTCTGGCAAATAATGTCACGACTAAAGCGAATGAG 2683

     101 ValValGlnThrGlnLeuArgGluGlnGlnAlaGluValGlyLysPhePh 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2684 GTTGTGCAAACCCAGTTACGCGAGCAGCAGGCAGAAGTCGGAAAGTTTTT 2733

     117 eAspIleSerGlyMetSerSerSerAlaValAlaLeuLeuAlaAlaAlaA 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2734 TGATATTAGCGGAATGTCTTCCAGTGCCGTTGCGCTGTTGGCTGCCGCGA 2783

     134 snThrLeuMetLeuThrLeuAsnGlnAlaAspSerLysLeuSerGlyLys 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2784 ATACGTTAATGCTGACGTTGAACCAGGCTGATAGCAAACTGTCTGGTAAG 2833

     151 LeuSerLeuValSerPheAspAlaAlaLysThrThrAlaSerSerMetMe 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2834 TTGTCATTAGTCAGTTTTGATGCAGCTAAAACGACGGCAAGCTCCATGAT 2883

     167 tArgGluGlyMetAsnAlaLeuSerGlySerIleSerGlnSerAlaLeuG 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2884 GCGCGAAGGGATGAATGCGTTGTCCGGTAGTATTTCCCAGAGCGCGCTTC 2933

     184 lnLeuGlyIleThrGlyValGlyAlaLysLeuGluTyrLysGlyLeuGln 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2934 AGTTGGGGATCACTGGCGTGGGCGCCAAACTGGAATATAAGGGGCTGCAG 2983

     201 AsnGluArgGlyAlaLeuLysHisAsnAlaAlaLysIleAspLysLeuTh 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2984 AATGAAAGAGGCGCGCTTAAACATAATGCCGCGAAGATCGATAAACTGAC 3033

     217 rThrGluSerHisSerIleLysAsnValLeuAsnGlyGlnAsnSerValL 234
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    3034 CACTGAAAGCCACAGTATTAAAAACGTGCTGAACGGGCAGAATAGCGTCA 3083

     234 ysLeuGlyAlaGluGlyValAspSerLeuLysSerLeuAsnMetLysLys 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    3084 AACTCGGTGCTGAAGGCGTCGATTCTCTGAAATCGTTAAATAT.AAGAAA 3132

     251 ThrGlyThrAspAlaThrLysAsnLeuAsnAspAlaThrLeuLysSerAs 267
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    3133 ACCGGTACCGATGCGACGAAAAATCTTAATGATGCGACGCTTAAATCTAA 3182

     267 nAlaGlyThrSerAlaThrGluSerLeuGlyIleLysAspSerAsnLysG 284
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    3183 TGCCGGAACCAGCGCCACGGAAAGTCTGGGTATTAAAGACAGTAATAAAC 3232

     284 lnIleSerProGluHisGlnAla.IleLeuSerLysArgLeuGluSerVa 300
         ||       |||||||||       ||||||||||||||||||||||||||
    3233 AAAGTCTCCCTGAACAT...CTATATCTTGTCGAAACGTCTTGAGTCTGT 3279

     300 lGluSerAspIleArgLeuGluGlnAsnThrMetAspMetThrArgIleA 317
         ||||||||||||||||||||||||||||||    ||||||:::||||||||||
    3280 CGAATCCGATATTCGTCTTGAGCAGAATTACATGGATATTACCCGAATCG 3329

     317 sp.AlaArgLysMetGlnMetThrGlyAspLeuIleMetLysAsnSerVa 333
         || ||||||||||||||||||||||||||||||||||||||||||||||||
    3330 ATAGCGCGCAAGATGCAGATGACGGGCGATCTGATTATGAAGAACTCGGT 3379

     333 lThrValGlyGlyIleAlaGlyAlaSerGlyGlnTyrAlaAlaThrGln. 349
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    3380 CACGGTCGGTGGTATTGCAGGGGCGTCCGGGCAGTACGCCGCTACTCAGG 3429

     350 .GluArgSerGluGlnGlnIleSerGlnValAsnAsnArgValAlaSerT 366
          :::   ||||||||||||||||||||||||||||||||||||||||||||
    3430 TAACGTTTCCGAGCAGCAAATTAGCCAGGTGAATAACCGGGTTGCCAGCA 3479

     366 hrAlaSerAspGluAlaArgGluSerSerArgLysSerThrSerLeuIle 382
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    3480 CCGCATCGGACGAAGCCCGTGAAAGTTCACGTAAATCGACCAGCCTGATT 3529

     383 GlnGluMetLeuLysThrMetGluSerIleAsnGlnSerLysAlaSerAl 399
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    3530 CAGGAAATGCTGAAAACAATGGAGAGCATTAACCAGTCGAAAGCATCCGC 3579

     399 aLeuAlaAlaIleAlaGlyAsnIleArgAla 409
         |||||||||||||||||||||||||||||||
    3580 ACTCGCTGCTATCGCAGGCAATATTCGCGCT 3610

seq_name: gb_bal:SDU66877

seq_documentation_block:
LOCUS       SDU66877     2995 bp    DNA             BCT       29-AUG-1996
DEFINITION  Salmonella dublin SicA (sicA) and SipB (sipB) genes, complete cds.
ACCESSION   U66877
```

```
  51 LeuSerProThrLeuSerAlaProAlaProGlyValLeuThrGlnThrPr 67
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2502 TTGAGTCCTACACTGAGCGCGCCTGCGCCAGGGGTATTAACGCAAACCCC 2551

  67 oGlyThrIleThrSerPheLeuLysAlaSerIleGlnAsnThrAspMetA 84
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2552 CGGAACGATCACGTCCTTTTTAAAAGCCAGTATTCAAAATACCGACATGA 2601

  84 snGlnAspLeuAsnAlaLeuAlaAsnAsnValThrThrLysAlaAsnGlu 100
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2602 ATCAGGATTTGAATGCCCTGGCAAATAATGTCACGACTAAAGCGAATGAG 2651

 101 ValValGlnThrGlnLeuArgGluGlnGlnAlaGluValGlyLysPhePh 117
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2652 GTTGTGCAAACCCAGTTACGCGAGCAGCAGGCAGAAGTCGGAAAGTTTTT 2701

 117 eAspIleSerGlyMetSerSerSerAlaValAlaLeuLeuAlaAlaAlaA 134
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2702 TGATATTAGCGGAATGTCTTCCAGTGCCGTTGCGCTGTTGGCTGCCGCGA 2751

 134 snThrLeuMetLeuThrLeuAsnGlnAlaAspSerLysLeuSerGlyLys 150
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2752 ATACGTTAATGCTGACGTTGAACCAGGCTGATAGCAAACTGTCTGGTAAG 2801

 151 LeuSerLeuValSerPheAspAlaAlaLysThrThrAlaSerSerMetMe 167
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2802 TTGTCATTAGTCAGTTTTGATGCAGCTAAAACGACGGCAAGCTCCATGAT 2851

 167 tArgGluGlyMetAsnAlaLeuSerGlySerIleSerGlnSerAlaLeuG 184
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2852 GCGCGAAGGGATGAATGCGTTGTCCGGTAGTATTTCCCAGAGCGCGCTTC 2901

 184 lnLeuGlyIleThrGlyValGlyAlaLysLeuGluTyrLysGlyLeuGln 200
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2902 AGTTGGGGATCACTGGCGTGGGCGCCAAACTGGAATATAAGGGGCTGCAG 2951

 201 AsnGluArgGlyAlaLeuLysHisAsnAlaAlaLysIleAspLysLeuTh 217
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2952 AATGAAAGAGGCGCGCTTAAACATAATGCCGCGAAGATCGATAAACTGAC 3001

 217 rThrGluSerHisSerIleLysAsnValLeuAsnGlyGlnAsnSerValL 234
     ||||||||||||||||||||||||||||||||||||||||||||||||||
3002 CACTGAAAGCCACAGTATTAAAAACGTGCTGAACGGGCAGAATAGCGTCA 3051

 234 ysLeuGlyAlaGluGlyValAspSerLeuLysSerLeuAsnMetLysLys 250
     ||||||||||||||||||||||||||||||||||||||||||||||||||
3052 AACTTGGTGCTGAAGGCGTCGATTCTCTGAAATCGTTAAATATGAAGAAA 3101

 251 ThrGlyThrAspAlaThrLysAsnLeuAsnAspAlaThrLeuLysSerAs 267
     ||||||||||||||||||||||||||||||||||||||||||||||||||
3102 ACCGGTACCGATGCGACGAAAAATCTTAATGATGCGACGCTTAAATCTAA 3151

 267 nAlaGlyThrSerAlaThrGluSerLeuGlyIleLysAspSerAsnLysG 284
     ||||||||||||||||||||||||||||||||||:::|||||||||||||
3152 TGCCGGAACCAGCGCCACGGAAAGTCTGGGTATTAAAAACAGTAATAAAC 3201

 284 lnIleSerProGluHisGlnAlaIleLeuSerLysArgLeuGluSerVal 300
     ||||||||||||||||||||||||||||||||||||||||||||||||||
3202 AAATCTCCCCTGAACATCAGGCTATTCTGTCGAAACGTCTTGAGTCTGTC 3251

 301 GluSerAspIleArgLeuGluGlnAsnThrMetAspMetThrArgIleAs 317
     ||||||||||||||||||||||||||||||||||||||||||||||||||
3252 GAATCCGATATTCGTCTTGAGCAGAATACCATGGATATGACCCGAATCGA 3301

 317 pAlaArgLysMetGlnMetThrGlyAspLeuIleMetLysAsnSerValT 334
     ||||||||||||||||||||||||||||||||||||||||||||||||||
3302 TGCGCGCAAGATGCAGATGACGGGCGATCTGATTATGAAGAACTCAGTCA 3351

 334 hrValGlyGlyIleAlaGlyAlaSerGlyGlnTyrAlaAlaThrGlnGlu 350
     |||||||||||||||||||||||  |||||||||||||||||||||||||
3352 CGGTCGGTGGTATTGCAGGGGCGTCCAGGCAGTACGCCGCTACTCAGGAA 3401

 351 ArgSerGluGlnGlnIleSerGlnValAsnAsnArgValAlaSerThrAl 367
     |||||||||||||||||||||||||||||||||||||||||||||||||
3402 CGTTCCGAGCAGCAAATTAGCCAGGTGAATAACCGGGTTGCCAGCACCGC 3451

 367 aSerAspGluAlaArgGluSerSerArgLysSerThrSerLeuIleGlnG 384
     ||||||||||||||||||||||||||||||||||||||||||||||||||
3452 ATCGGACGAAGCCCGTGAAAGTTCACGTAAATCGACCAGCCTGATTCAGG 3501

 384 luMetLeuLysThrMetGluSerIleAsnGlnSerLysAlaSerAlaLeu 400
     ||||||||||||||||||||||||||||||||||||||||||||||||||
3502 AAATGCTGAAAACAATGGAGAGCATTAACCAGTCGAAAGCATCCGCACTC 3551

 401 AlaAlaIleAlaGlyAsnIleArgAla 409
     |||||||||||||||||||||||||
3552 GCTGCTATCGCAGGCAATATTCGCGCT 3578
```

seq_name: gb_ba1:STU25631

seq_documentation_block:

```
LOCUS       STU25631     3622 bp    DNA             BCT       06-SEP-1995
DEFINITION  Salmonella typhimurium invasin homologs SicA (sicA), SipB (sipB)
            and SipC  (sipC) genes, complete cds.
ACCESSION   U25631
NID         g975293
VERSION     U25631.1  GI:975293
KEYWORDS    .
SOURCE      Salmonella typhimurium.
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Salmonella.
REFERENCE   1  (bases 1 to 3622)
  AUTHORS   Kaniga,K., Tucker,S., Trollinger,D. and Galan,J.E.
  TITLE     Homologs of the Shigella IpaB and IpaC invasins are required for
            Salmonella typhimurium entry into cultured epithelial cells
  JOURNAL   J. Bacteriol. 177 (14), 3965-3971 (1995)
  MEDLINE   95332200
REFERENCE   2  (bases 1 to 3622)
  AUTHORS   Galan,J.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-APR-1995) Jorge E. Galan, Molecular Genetics and
            Microbiology, School of Medicine, SUNY Stony Brook, Stony Brook, NY
            11794-5222, USA
FEATURES             Location/Qualifiers
     source          1. .3622
                     /organism="Salmonella typhimurium"
                     /strain="SL1344"
                     /db_xref="taxon:602"
     gene            75. .572
                     /gene="sicA"
     CDS             75. .572
                     /gene="sicA"
                     /codon_start=1
                     /transl_table=11
                     /product="SicA"
                     /protein_id="AAA75168.1"
                     /db_xref="PID:g975294"
                     /db_xref="GI:975294"
                     /translation="MDYQNNVSEERVAEMIWDAVSEGATLKDVHGIPQDMMDGLYAHA
                     YEFYNQGRLDEAETFFRFLCIYDFYNPDYTMGLAAVCQLKKQFQKACDLYAVAFTLLK
                     NDYRPVFFTGQCQLLMRKAAKARQCFELVNERTEDESLRAKALVYLEALKTAETEQHS
                     EQEKE"
     gene            575. .2356
                     /gene="sipB"
     CDS             575. .2356
                     /gene="sipB"
                     /codon_start=1
                     /transl_table=11
                     /product="SipB"
                     /protein_id="AAA75169.1"
                     /db_xref="PID:g975295"
                     /db_xref="GI:975295"
                     /translation="MVNDASSISRSGYTQNPRLAEAAFEGVRKNTDFLKAADKAFKDV
                     VATKAGDLKAGTKSGESAINTVGLKPPTDAAREKLSSEGQLTLLLGKLMTLLGDVSLS
                     QLESRLAVWQAMIESQKEMGIQVSKEFQTALGEAQEATDLYEASIKKTDTAKSVYDAA
```

```
     401 AlaAlaIleAlaGlyAsnIleArgAla 409
         ||||||||||||||||||||||||||
    2098 GCTGCTATCGCAGGCAATATTCGCGCT 2124

seq_name: gb_ba1:STSIPGNS

seq_documentation_block:
LOCUS       STSIPGNS     5393 bp    DNA             BCT       04-OCT-1995
DEFINITION  S.typhi spaT, sipB, sipC, sipD and sipA genes.
ACCESSION   X82670
NID         g1009001
VERSION     X82670.1  GI:1009001
KEYWORDS    sipA gene; sipB gene; sipC gene; sipD gene; spaT gene.
SOURCE      Salmonella typhi.
  ORGANISM  Salmonella typhi
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Salmonella.
REFERENCE   1  (bases 1 to 5393)
  AUTHORS   Hermant,D., Menard,R., Arricau,N., Parsot,C. and Popoff,M.Y.
  TITLE     Functional conservation of the Salmonella and Shigella effectors of
            entry into epithelial cells
  JOURNAL   Mol. Microbiol. 17 (4), 781-789 (1995)
  MEDLINE   96111497
REFERENCE   2  (bases 1 to 5393)
  AUTHORS   Popoff,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-1994) M. Popoff, Institut Pasteur, Unite des
            Enterobacteries, 28 rue du Docteur Roux, 75724 Paris Cedex 15,
            FRANCE
FEATURES             Location/Qualifiers
     source          1. .5393
                     /organism="Salmonella typhi"
                     /strain="Ty2"
                     /db_xref="taxon:601"
     gene            43. .540
                     /gene="spaT"
     CDS             43. .540
                     /gene="spaT"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA57987.1"
                     /db_xref="PID:g1009002"
                     /db_xref="GI:1009002"
                     /db_xref="SPTREMBL:Q57260"
                     /translation="MDYQNNVSEERVAEMIWDAVSEGATLKDVHGIPQDMMDGLYAHA
                     YEFYNQGRLDEAETFFRFLCIYDFYNPDYTMGLAAVCQLKKQFQKACDLYAVAFTLLK
                     NDYRPVFFTGQCQLLMRKAAKARQCFELVNERTEDESLRAKALVYLEALKTAETEQHS
                     EQEKE"
     gene            543. .2324
                     /gene="sipB"
     CDS             543. .2324
                     /gene="sipB"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA57988.1"
                     /db_xref="PID:g1009003"
                     /db_xref="GI:1009003"
                     /db_xref="SPTREMBL:Q56134"
                     /translation="MVNDASSISRSGYTQNPRLAEAAFEGVRKNTDFLKAADKAFKDV
                     VAIKAGDLKAGTKSGESAINTVGLKPPTDAAREKLSSEGQLTLLLGKLMTLLGDVSLS
                     QLESRLAVWQAMIESQKEMGIQVSKEFQTALGEAQEATDLYEASIKKTDTAKSVYDAA
                     AKKLTQAQNKLQSLDPADPGYAQAEAAVEQAGKEATEAKEALDKATDATVKAGTDAKA
                     KAEKADNILTKFQGTANAASQNQVSQGEQDNLSNVARLTMLMAMFIEIVGKNTEESLQ
                     NDLALFNALQEGRQAEMEKKSAEFQEETRKAEETNRIMGCIGKVLGALLTIVSVVAAV
                     FTGGASLALAAVGLAVMVADEIVKAATGVSFIQQALNPIMEHVLKPLMELIGKAITKA
                     LEGLGVDKKTAEMAGSIVGAIVAAIAMVAVIVVVAVVGKGAAAKLGNALSKMMGETIK
                     KLVPNVLKQLAQNGSKLFTQGMQRITSGLGNVGSKMGLQTNALSKELVGNTLNKVALG
                     MEVTNTAAQSAGGVAEGVFIKNASEALADFMLARFAMDQIQQWLKQSVEIFGENQKVT
                     AELQKAMSSAVQQNADASRFILRQSRA"
     gene            2352. .3581
                     /gene="sipC"
     CDS             2352. .3581
                     /gene="sipC"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA57989.1"
                     /db_xref="PID:g1009004"
                     /db_xref="GI:1009004"
                     /db_xref="SPTREMBL:Q56135"
                     /translation="MLISNVGINPAAYLNNHSVENSSQTASQSVSAKDILNSIGISSS
                     KVSDLGLSPTLSAPAPGVLTQTPGTITSFLKASIQNTDMNQDLNALANNVTTKANEVV
                     QTQLREQQAEVGKFFDISGMSSSAVALLAAANTLMLTLNQADSKLSGKLSLVSFDAAK
                     TTASSMMREGMNALSGSISQSALQLGITGVGAKLEYKGLQNERGALKHNAAKIDKLTT
                     ESHSIKNVLNGQNSVKLGAEGVDSLKSLNMKKTGTDATKNLNDATLKSNAGTSATESL
                     GIKNSNKQISPEHQAILSKRLESVESDIRLEQNTMDMTRIDARKMQMTGDLIMKNSVT
                     VGGIAGASRQYAATQERSEQQISQVNNRVASTASDEARESSRKSTSLIQEMLKTMESI
                     NQSKASALAAIAGNIRA"
     gene            3651. .4670
                     /gene="sipD"
     CDS             3651. .4670
                     /gene="sipD"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA57990.1"
                     /db_xref="PID:g1009005"
                     /db_xref="GI:1009005"
                     /db_xref="SPTREMBL:Q56136"
                     /translation="MLNIQNYSASPHPGIVAERPQTPSASEHAEIAVVPSTTEHRGTD
                     IISLSQAATKIQQAQQTLQSTPPISEENNDERTLARQQLTSSLNALAKSGVSLSAEQN
                     ENLRSTFSARRRPYLALRLWPARTTISDAEIWDMVSQNISAIGDSYLGVYENVVAVYT
                     DFYQAFSDILSKMGGWLSPGKDGNTIKLNVDSLKSEISSLINKYTQINKNTILFPSQT
                     GSGMTTATKAEAEQWIKELNLPDSCLKASGSGYVVLVDTGPLSKMVSDLNGIGSGSAL
                     ELDNAKYQAWQSGFKAQEENLKTTLQTLTQKYSNANSLYDNLVKVLSSTISSSLETAK
                     SFLQG"
     gene            4689. .5393
                     /gene="sipA"
     CDS             4689. .>5393
                     /gene="sipA"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA57991.1"
                     /db_xref="PID:g1009006"
                     /db_xref="GI:1009006"
                     /db_xref="SPTREMBL:Q56137"
                     /translation="MVTSVRTQPPVIMPGMQTEIKTQATNLAANLSAVRESATATLSG
                     EIKGQQLEDFPALIKQASLDALFKCGKDAEALKEVFTNSNNVAGKKAIMEFAGLFRSA
                     LNATSDSPEAKTLLMKVGAEYTAQIIKDGLKEKSAFGPWLPETKKAEAKLENLEKQLL
                     DIIKNNTGGELSKLSTNLVMQEVMPYIASCIEHNFGCTLDPLTRSSLTQLVDKAAAKA
                     VEALDMCHQKLTQEQGT"
BASE COUNT     1584 a    1156 c    1458 g    1195 t
ORIGIN


alignment_scores:
            Quality: 1640.00              Length:    409
              Ratio:  4.020                Gaps:      0
 Percent Similarity:  99.756   Percent Identity: 99.511

alignment_block:
 US-09-068-804-6 x STSIPGNS   ..

 Align seg 1/1  to: STSIPGNS  from: 1  to: 5393

       1 MetLeuIleSerAsnValGlyIleAsnProAlaAlaTyrLeuAsnAsnHi 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2352 ATGTTAATTAGTAATGTGGGAATAAATCCCGCCGCTTATTTAAATAATCA 2401

      17 sSerValGluAsnSerSerGlnThrAlaSerGlnSerValSerAlaLysA 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2402 TTCTGTTGAGAATAGTTCACAGACAGCTTCGCAATCCGTTAGCGCTAAAG 2451

      34 spIleLeuAsnSerIleGlyIleSerSerSerLysValSerAspLeuGly 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2452 ATATTCTGAATAGTATTGGTATTAGCAGCAGTAAAGTCAGTGACCTGGGG 2501
```

```
                GIKDSNKQISPEHQAILSKRLESVESDIRLEQNTMDMTRIDARKMQMTGDLIMKNSVT
                VGGIAGASGQYAATQERSEQQISQVNNRVASTASDEARESSRKSTSLIQEMLKTMESI
                NQSKASALAAIAGNIRA"
     gene       2198. .3229
                /gene="sspD"
     CDS        2198. .3229
                /gene="sspD"
                /note="secreted protein"
                /codon_start=1
                /transl_table=11
                /function="invasion determinant"
                /evidence=experimental
                /product="SspD"
                /protein_id="AAC43548.1"
                /db_xref="PID:g1117814"
                /db_xref="GI:1117814"
                /translation="MLNIQNYSASPHPGIVAERPQTPSASEHVETAVVPSTTEHRGTD
                IISLSQAATKIHQAQQTLQSTPPISEENNDERTLARQQLTSSLNALAKSGVSLSAEQN
                ENLRSAFSAPTSALFSASPMAQPRTTISDAEIWDMVSQNISAIGDSYLGVYENVVAVY
                TDFYQAFSDILSKMGGWLLPGKDGNTVKLDVTSLKNDLNSLVNKYNQINSNTVLFPAQ
                SGSGVKVATEAEARQWLSELNLPNSCLKSYGSGYVVTVDLTPLQKMVQDIDGLGAPGK
                DSKLEMDNAKYQAWQSGFKAQEENMKTTLQTLTQKYSNANSLYDNLVKVLSSTISSSL
                ETAKSFLQG"
     gene       3248. .3441
                /gene="sspA"
     CDS        3248. .>3441
                /gene="sspA"
                /note="secreted protein"
                /codon_start=1
                /transl_table=11
                /function="unknown"
                /product="SspA"
                /protein_id="AAC43549.1"
                /db_xref="PID:g1117815"
                /db_xref="GI:1117815"
                /translation="MVTSVRTQPPVIMPGMQTEIKTQATNLAANLSAVRESATATLSG
                EIKGPQLEDFPALIKQASLDA"
BASE COUNT      978 a    759 c    925 g    779 t
ORIGIN


alignment_scores:
           Quality: 1649.00            Length:    409
             Ratio:  4.032               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-068-804-6 x STU30491   ..

 Align seg 1/1  to: STU30491  from: 1  to: 3441

       1 MetLeuIleSerAsnValGlyIleAsnProAlaAlaTyrLeuAsnAsnHi 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     898 ATGTTAATTAGTAATGTGGGAATAAATCCCGCCGCTTATTTAAATAATCA 947

      17 sSerValGluAsnSerSerGlnThrAlaSerGlnSerValSerAlaLysA 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     948 TTCTGTTGAGAATAGTTCACAGACAGCTTCGCAATCCGTTAGCGCTAAAG 997

      34 spIleLeuAsnSerIleGlyIleSerSerSerLysValSerAspLeuGly 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     998 ATATTCTGAATAGTATTGGTATTAGCAGCAGTAAAGTCAGTGACCTGGGG 1047

      51 LeuSerProThrLeuSerAlaProAlaProGlyValLeuThrGlnThrPr 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1048 TTGAGTCCTACACTGAGCGCGCCTGCGCCAGGGGTATTAACGCAAACCCC 1097

      67 oGlyThrIleThrSerPheLeuLysAlaSerIleGlnAsnThrAspMetA 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1098 CGGAACGATCACGTCCTTTTTAAAAGCCAGTATTCAAAATACCGACATGA 1147

      84 snGlnAspLeuAsnAlaLeuAlaAsnAsnValThrThrLysAlaAsnGlu 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||

    1148 ATCAGGATTTGAATGCTCTGGCAAATAATGTCACGACTAAAGCGAATGAG 1197

     101 ValValGlnThrGlnLeuArgGluGlnGlnAlaGluValGlyLysPhePh 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1198 GTTGTGCAAACCCAGTTACGCGAGCAGCAGGCAGAAGTCGGAAAGTTTTT 1247

     117 eAspIleSerGlyMetSerSerSerAlaValAlaLeuLeuAlaAlaAlaA 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1248 TGATATTAGCGGAATGTCTTCCAGTGCCGTTGCGCTGTTGGCTGCCGCGA 1297

     134 snThrLeuMetLeuThrLeuAsnGlnAlaAspSerLysLeuSerGlyLys 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1298 ATACGTTAATGCTGACGTTGAACCAGGCTGATAGCAAACTGTCTGGTAAG 1347

     151 LeuSerLeuValSerPheAspAlaAlaLysThrThrAlaSerSerMetMe 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1348 TTGTCATTAGTCAGTTTTGATGCAGCTAAAACGACGGCAAGCTCCATGAT 1397

     167 tArgGluGlyMetAsnAlaLeuSerGlySerIleSerGlnSerAlaLeuG 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1398 GCGCGAAGGGATGAATGCGTTGTCCGGTAGTATTTCCCAGAGCGCGCTTC 1447

     184 lnLeuGlyIleThrGlyValGlyAlaLysLeuGluTyrLysGlyLeuGln 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1448 AGTTGGGGATCACTGGCGTGGGCGCCAAACTGGAATATAAGGGGCTGCAG 1497

     201 AsnGluArgGlyAlaLeuLysHisAsnAlaAlaLysIleAspLysLeuTh 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1498 AATGAAAGAGGCGCGCTTAAACATAATGCCGCGAAGATCGATAAACTGAC 1547

     217 rThrGluSerHisSerIleLysAsnValLeuAsnGlyGlnAsnSerValL 234
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1548 CACTGAAAGCCACAGTATTAAAAACGTGCTGAACGGGCAGAATAGCGTCA 1597

     234 ysLeuGlyAlaGluGlyValAspSerLeuLysSerLeuAsnMetLysLys 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1598 AACTCGGTGCTGAAGGCGTCGATTCTCTGAAATCGTTAAATATGAAGAAA 1647

     251 ThrGlyThrAspAlaThrLysAsnLeuAsnAspAlaThrLeuLysSerAs 267
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1648 ACCGGTACCGATGCGACGAAAAATCTTAATGATGCGACGCTTAAATCTAA 1697

     267 nAlaGlyThrSerAlaThrGluSerLeuGlyIleLysAspSerAsnLysG 284
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1698 TGCCGGAACCAGCGCCACGGAAAGTCTGGGTATTAAAGACAGTAATAAAC 1747

     284 lnIleSerProGluHisGlnAlaIleLeuSerLysArgLeuGluSerVal 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1748 AAATCTCCCCTGAACATCAGGCTATTCTGTCGAAACGTCTTGAGTCTGTC 1797

     301 GluSerAspIleArgLeuGluGlnAsnThrMetAspMetThrArgIleAs 317
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1798 GAATCCGATATTCGTCTTGAGCAGAATACCATGGATATGACCCGAATCGA 1847

     317 pAlaArgLysMetGlnMetThrGlyAspLeuIleMetLysAsnSerValT 334
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1848 TGCGCGCAAGATGCAGATGACGGGCGATCTGATTATGAAGAACTCGGTCA 1897

     334 hrValGlyGlyIleAlaGlyAlaSerGlyGlnTyrAlaAlaThrGlnGlu 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1898 CGGTCGGTGGTATTGCAGGGGCGTCCGGGCAGTACGCCGCTACTCAGGAA 1947

     351 ArgSerGluGlnGlnIleSerGlnValAsnAsnArgValAlaSerThrAl 367
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1948 CGTTCCGAGCAGCAAATTAGCCAGGTGAATAACCGGGTTGCCAGCACCGC 1997

     367 aSerAspGluAlaArgGluSerSerArgLysSerThrSerLeuIleGlnG 384
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1998 ATCGGACGAAGCCCGTGAAAGTTCACGTAAATCGACCAGCCTGATTCAGG 2047

     384 luMetLeuLysThrMetGluSerIleAsnGlnSerLysAlaSerAlaLeu 400
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2048 AAATGCTGAAAACAATGGAGAGCATTAACCAGTCGAAAGCATCCGCACTC 2097
```

OM of: US-09-068-804-6 to: GenEmbl:* out_format : pfs

Date: Oct 5, 1999 7:59 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09068804/runat_05101999_145134_10315/app_query.fasta.1
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=pam150 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=escore -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfs -NORM=stat -USER=US09068804 -NCPU=6 -ICPU=3 -WAIT
-THREADS=1

Search information block:
Query: US-09-068-804-6
Query length: 410
Database: GenEmbl:*
Database sequences: 679419
Database length: 1590154680
Search time (sec): 1424.920000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ..
gb_ba2:STU30491	+	1649.00	2349.97	5.9e-123	3441	! U30491 Salmonella typhimurium
gb_ba1:STSIPGNS	+	1640.00	2333.42	4.9e-122	5393	! X82670 S.typhi spaT, sipB, sip
gb_ba1:STU25631	+	1506.50	2145.64	1.4e-111	3622	! U25631 Salmonella typhimurium
gb_ba1:SDU66877	+	739.00	1048.92	1.7e-50	2995	! U66877 Salmonella dublin SicA
gb_ba1:W10PPIBCA	+	380.50	532.13	1.1e-21	4761	! M34849 Plasmid pWR100 (from S.
gb_ba1:SHFIPABCD	+	380.50	531.32	1.2e-21	5255	! J04117 S.flexneri ipaBCD regio
gb_ba1:SFPVR2	+	379.00	526.10	2.3e-21	7663	! X15319 Shigella flexineri plas
gb_ba1:SDIPABCD	+	377.50	524.84	2.7e-21	6869	! X60777 Shigella dysenteriae ip
gb_pl1:YSCINTANA	+	113.00	151.34	1.73	3728	! L03188 Saccharomyces cerevisia
gb_pl1:SCCPN60	+	108.50	151.28	1.74	1707	! Z68903 S.cereale cv. Petkus 'P
gb_pl1:SCYDL057W	+	115.00	150.81	1.85	5648	! Z74105 S.cerevisiae chromosome
gb_pl1:SCUSO1	+	115.00	150.49	1.93	5877	! X54378 S.cerevisiae gene for U
gb_pl1:SCYDL058W	+	115.00	149.69	2.13	6480	! Z74106 S.cerevisiae chromosome
gb_ov:GDVMHC	+	110.00	148.13	2.61	3263	! X59552 G.domesticus mRNA for v
gb_pat:I11729	+	109.50	147.28	2.91	3319	! I11729 Sequence 2 from Patent
gb_pat:I88957	+	109.50	147.28	2.91	3319	! I88957 Sequence 1 from patent
gb_in1:LEIKINLIKE	+	109.50	147.28	2.91	3319	! L07879 Leishmania chagasi kine
gb_ov:GGA16534	+	103.50	147.01	3.01	1199	! Y16534 Gallus gallus mRNA for
gb_ba1:AF026811	+	108.00	146.67	3.15	2751	! AF026811 Butyrivibrio fibrisol
gb_pr4:AF111782	+	110.50	144.01	4.42	5902	! AF111782 Homo sapiens extraocu
gb_pl1:PSU21139	+	104.50	143.63	4.65	2162	! U21139 Pisum sativum chaperoni
gb_ba1:STREMM24	+	104.00	143.03	5.02	2131	! M19031 S.pyogenes type 24 M-pr
gb_ba1:MHLMP34GN	+	110.50	142.86	5.13	6797	! X95601 M.hominis lmp3 and lmp4
gb_ba2:ECOTRG	+	102.50	142.58	5.31	1731	! K02073 Escherichia coli Trg se
gb_pr2:HUMEEAP	+	108.50	142.46	5.40	5028	! L40157 Human endosome-associat
gb_om:OCHEAMYO1	+	105.50	142.08	5.67	3114	! Z34886 O.cuniculus mRNA for be
gb_ba2:AF034768	+	100.50	141.25	6.30	1434	! AF034768 Pseudomonas stutzeri
gb_pl2:STU46136	+	102.50	141.09	6.43	2078	! U46136 Solanum tuberosum chape
gb_ba1:ECAJ2736	+	103.00	138.95	8.46	2948	! AJ002736 Enterobacter cloacae
gb_pr3:HSU22815	+	104.50	138.65	8.79	3976	! U22815 Human LAR-interacting p
gb_pr2:HSBMHCV	+	103.50	138.10	9.44	3570	! X51591 Human beta-MHC mRNA for
gb_pr3:HSU22816	+	104.50	137.99	9.57	4313	! U22816 Human LAR-interacting p
gb_ov:AF014012	+	104.50	137.39	10.33	4638	! AF014012 Gallus gallus restin
gb_ba1:HECFLAGELB	+	99.00	137.25	10.52	1800	! L08908 Helicobacter mustelae f
gb_pr2:HSP162	+	104.50	137.15	10.66	4780	! X78998 H.sapiens p162 mRNA. 1/
gb_pr1:D78367	+	99.00	136.94	10.96	1871	! D78367 Human mRNA for K12 kera
gb_ba1:PAFGK979	+	99.00	136.00	12.36	2099	! X98465 P.aeruginosa flagellin
gb_pl1:SS95BASAL	+	101.00	135.64	12.94	3113	! AJ001438 Spermatozopsis simili
gb_in1:PMU59294	+	104.50	135.53	13.12	5826	! U59294 Placopecten magellanicu
gb_in1:PMU59295	+	104.50	135.49	13.18	5853	! U59295 Placopecten magellanicu
gb_ro:RNMHCB	+	104.50	135.39	13.35	5925	! X15939 Rat mRNA for beta cardi
gb_pr1:AB007914	+	105.00	135.39	13.36	6471	! AB007914 Homo sapiens mRNA for
gb_ba1:KOAJ2735	+	100.50	135.38	13.38	2945	! AJ002735 Klebsiella oxytoca
gb_om:SSU75316	+	104.50	135.30	13.52	5996	! U75316 Sus scrofa beta-myos
gb_ba2:AE000072	-	108.00	134.94	14.16	11572	! AE000072 Rhizobium sp. NGR

seq_name: gb_ba2:STU30491

seq_documentation_block:
LOCUS STU30491 3441 bp DNA BCT 19-JAN-1996
DEFINITION Salmonella typhimurium SspA (sspA) and SspB (sspB) genes, partial cds, and SspC (sspC) and SspD (sspD) genes, complete cds.
ACCESSION U30491
NID g1117811
VERSION U30491.1 GI:1117811
KEYWORDS .
SOURCE Salmonella typhimurium.
ORGANISM Salmonella typhimurium
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Salmonella.
REFERENCE 1 (bases 1 to 3441)
AUTHORS Hueck,C.J., Hantman,M.J., Bajaj,V., Johnston,C., Lee,C.A. and Miller,S.I.
TITLE Salmonella typhimurium secreted invasion determinants are homologous to Shigella Ipa proteins
JOURNAL Mol. Microbiol. 18 (3), 479-490 (1995)
MEDLINE 96342378
REFERENCE 2 (bases 1 to 3441)
AUTHORS Miller,S.I.
TITLE Direct Submission
JOURNAL Submitted (28-JUN-1995) Samuel I. Miller, Infectious Disease Unit, Massachusetts General Hospital, Harvard Medical School, Fruit Street, Boston, MA 02114, USA
FEATURES Location/Qualifiers
source 1. .3441
/organism="Salmonella typhimurium"
/strain="SL1344"
/db_xref="taxon:602"
/map="59 min"
gene 1. .870
/gene="sspB"
CDS <1. .870
/gene="sspB"
/note="secreted protein"
/codon_start=1
/transl_table=11
/function="invasion determinant"
/product="SspB"
/protein_id="AAC43546.1"
/db_xref="PID:g1117812"
/db_xref="GI:1117812"
/translation="RKAEETNRIMGCIGKVLGALLTIVSVVAAVFTGGASLALAAVGL
AVMVADEIVKAATGVSFIQQALNPIMEHVLKPLMELIGKAITKALEGLGVDKKTAEMA
GSIVGAIVAAIAMVAVIVVVAVVGKGAAAKLGNALSKMMGETIKKLVPNVLKQLAQNG
SKLFTQGMQRITSGLGNVGSKMGLQTNALSKELVGNTLNKVALGMEVTNTAAQSAGGV
AEGVFIKNASEALADFMLARFAMDQIQQWLKQSVEIFGENQKVTAELQKAMSSAVQQN
ADASRFILRQSRA"
gene 898. .2127
/gene="sspC"
CDS 898. .2127
/gene="sspC"
/note="secreted protein"
/codon_start=1
/transl_table=11
/function="invasion determinant"
/evidence=experimental
/product="SspC"
/protein_id="AAC43547.1"
/db_xref="PID:g1117813"
/db_xref="GI:1117813"
/translation="MLISNVGINPAAYLNNHSVENSSQTASQSVSAKDILNSIGISSS
KVSDLGLSPTLSAPAPGVLTQTPGTITSFLKASIQNTDMNQDLNALANNVTTKANEVV
QTQLREQQAEVGKFFDISGMSSSAVALLAAANTLMLTLNQADSKLSGKLSLVSFDAAK
TTASSMMREGMNALSGSISQSALQLGITGVGAKLEYKGLQNERGALKHNAAKIDKLTT
ESHSIKNVLNGQNSVKLGAEGVDSLKSLNMKKTGTDATKNLNDATLKSNAGTSATESL

:::::::::
4617 TTAAATATC 4625

```
          :::::::::
    4567 TTAAATATC 4575

seq_name: N_Geneseq_36:Q49508

seq_documentation_block:
ID   Q49508 standard; DNA; 9220 BP.
AC   Q49508;
DT   26-APR-1994  (first entry)
DE   Gene cluster for high molecular weight protein 1 (HMW1).
KW   HMW; high molecular weight protein; virus; vaccine; influenza;
KW   epitope; immunity; haemophilus influenzae; gene cluster; ss.
OS   Haemophilus influenzae.
FH   Key             Location/Qualifiers
FT   cds             351. .4958
FT                   /*tag=  a
FT                   /product= High molecular weight protein 1.
FT   cds             5114. .6748
FT                   /*tag=  b
FT                   /note= "One of a gene cluster for high molecular
FT                   weight protein 1"
FT   cds             7062. .9011
FT                   /*tag=  c
FT                   /note= "One of a gene cluster for high molecular
FT                   weight protein 1"
FT   repeat_unit     2351. .2380
FT                   /*tag=  d
FT                   /label=
FT                   /note= "Possible error in sequence. Alternative
FT                   sequence for this region is CTTAAATGTT
FT                   TCCGAGAGTG GCGAGTTTAAA (See R41725)"
PN   WO9319090-A.
PD   30-SEP-1993.
PF   16-MAR-1993; U02166.
PR   16-MAR-1992; GB-005704.
PA   (BARE/) BARENKAMP S J.
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PI   Barenkamp SJ;
DR   WPI; 93-320683/40.
DR   P-PSDB; R41725, R41726, R41727.
PT   High molecular weight surface proteins - of non-typeable
PT   haemophilus which exhibit immunogenic properties
PS   Claim 9; Figure 6; 100pp; English.
CC   The isolation and purification of the high molecular weight protein
CC   enables the identification of the major protective epitopes of the
CC   protein by conventional epitope mapping.  These epitopes can then be
CC   synthesised using standard techniques and incorporated into fully
CC   synthetic or recombinant vaccines.
SQ   Sequence   9220 BP;    3101 A;    1861 C;    1805 G;    2453 T;


alignment_scores:
            Quality:  102.50              Length:    353
              Ratio:  0.483                Gaps:     15
 Percent Similarity:  60.057   Percent Identity: 19.547

alignment_block:
 US-09-068-804-6 x Q49508   ..

 Align seq 1/1  to: Q49508  from: 1  to: 9220

      15 AsnAsnHisSerValGluAsnSerSerGlnThrAlaSerGlnSerValSe 31
         ||||||:::::::   ::::::|||:::::::::::|||:::   ::::::::
    3666 AATAACAACACTGAAGATAGCAGTGACAATAATGCCGGCTTAACTATCGA 3715

      31 rAlaLysAspIleLeuAsnSerIleGlyIleSerSerSerLysValSerA 48
         :||||||::::::      :::   :::|||:::|||   |||:::   :
    3716 TGCAAAAAATGTAACAGTAAACAACAATATTACTTCTCACAAAGCAGTGA 3765

      48 spLeuGly..................................... 50
         ::::::::
    3766 GCATCTCTGCGACAAGTGGAGAAATTACCACTAAAACAGGTACAACCATT 3815

      51 .....................................Le 51
                                               ||
    3816 AACGCAACCACTGGTAACGTGGAGATAACCGCTCAAACAGGTAGTATCCT 3865

      51 uSerProThrLeuSerAlaProAlaProGlyValLeuThrGlnThrProG 68
         |:::   :::   |||:::::::::::::   :::||||||   |||   |
    3866 AGGTGGAATTGAGTCCAGCTCTGGCTCTGTAACACTTACTGCAACCGAGG 3915

      68 ly......ThrIleThrSerPheLeuLysAlaSerIleGlnAsnThrAsp 82
         ||       ::::::::::::::::     :::::::::      |||:::
    3916 GCGCTCTTGCTGTAAGCAATATTTCGGGCAACACCGTTACTGTTACTGCA 3965

      83 MetAsnGlnAspLeuAsnAlaLeuAlaAsnAsnValThrThrLysAlaAs 99
         :::   :::|||:::::::||||||::::::   |||:::|||:::::
    3966 AATAGCGGTGCATTAACCACTTTGGCAGGCTCT...ACAATTAAAGGAAC 4012

      99 nGluValValGlnThrGlnLeuArgGluGlnGlnAlaGluValGlyLysP 116
         :|||   |||   |||            |||   ::::::::::|||
    4013 CGAGAGTGTAACCACT......TCAAGTCAATCAGGCGATATCGGC.... 4052

     116 hePheAspIleSerGlyMetSerSerSerAlaValAlaLeuLeuAlaAla 132
                 :::   :::|||::::::::::|||:::::::   |||:::
    4053 ..........GGTACGATTTCTGGTGGCACAGTAGAGGTTAAAGCAACC 4091

     133 AlaAsnThrLeuMetLeuThrLeuAsnGlnAlaAspSerLysLeuSerGl 149
         ::::::         ||||||   :::|||::::::||||||:::   ::
    4092 GAAAGT.........TTAACC...ACTCAATCCAATTCAAAAATTAAAGC 4129

     149 yLysLeuSerLeuValSerPheAspAlaAlaLysThrThrAlaSerSerM 166
         ::::   :::             :::|||:::   |||:::|||::::::
    4130 AACAACAGGC............GAGGCTAACGTAACAAGTGCAACAGGTA 4167

     166 etMetArgGluGlyMetAsnAlaLeuSerGlySerIleSerGlnSerAla 182
           :::          :::::::::   :::   :::::::::   :::|||
    4168 CAATTGGTGGTACGATTTCCGGTAATACGGTAAATGTTACGGCAAACGCT 4217

     183 LeuGlnLeuGlyIleThrGlyValGlyAlaLysLeuGluTyrLysGlyLe 199
         :::|||   :::   |||   ||||||   ::::::
    4218 GGCGATTTAACAGTT...GGGAATGGCGCAGAAATTAAT.........GC 4255

     199 uGlnAsnGluArgGlyAlaLeuLysHisAsnAlaAlaLysIleAspLysL 216
           :::       ::::::|||:::   :::::::::          ||||
    4256 GACAGAAGGAGCTGCAACCTTAACTACATCATCGGGC.........AAAT 4296

     216 euThrThrGluSerHisSerIleLysAsnValLeuAsnGlyGlnAsnSer 232
         ||||||||||||:::   |||       ::::::   ::::::::::::
    4297 TAACTACCGAAGCTAGTTCA......CACATTACTTCAGCCAAGGGTCAG 4340

     233 ValLysLeuGlyAlaGluGlyValAspSerLeuLysSerLeuAsnMetLy 249
         |||:::|||:::|||::::::   :::          |||:::|||
    4341 GTAAATCTTTCAGCTCAGGATGGTAGCGTTGCAGGAAGTATTAATGCCGC 4390

     249 sLysThrGlyThrAspAlaThrLysAsnLeuAsnAspAlaThrLeuLysS 266
          ::::::      ::::::|||   :::|||      :::|||:::|||:
    4391 CAATGTGACACTAAATACTACAGGCACTTTA......ACTACCGTGAAGG 4434

     266 erAsnAlaGlyThrSerAlaThrGluSerLeuGlyIleLysAspSerAsn 282
         ::::::::   ::::::::::::   :::|||   |||::::::   :::
    4435 GTTCAAACATTAATGCAACCAGCGGTACCTTGGTTATTAACGCAAAAGAC 4484

     283 LysGlnIleSerProGluHisGlnAlaIleLeuSerLysArgLeuGluSe 299
         ::::::         :::   |||   |||::::::::::
    4485 GCTGAGCTA.........AATGGCGCAGCATTGGGTAACCACACAGTGGT 4525

     299 rValGluSerAspIleArgLeuGluGlnAsnThrMetAspMetThrArgI 316
           :::::::::                :::::::::::   |||
    4526 AAATGCAACCAACGCAAATGGCTCCGGCAGCGTAATCGCGACAACC.... 4571

     316 leAspAlaArgLysMetGlnMetThrGlyAspLeuIleMetLysAsnSer 332
           :::   ::::::::::::|||||||||||||||      |||:::
    4572 .....TCAAGCAGAGTGAACATCACTGGGGATTTAATCACAATAAATGGA 4616

     333 ValThrVal 335
```

```
     391 erIleAsnGlnSerLysAlaSerAlaLeuAla 401
         ::   |||   :::|||::::::|||::::::
    6322 CGAGTAACGGAGATAAAGGGGCAGCAATGAGT 6353

seq_name: N_Geneseq_36:T90996

seq_documentation_block:
ID   T90996 standard; DNA; 9171 BP.
AC   T90996;
DT   14-APR-1998  (first entry)
DE   Non-typeable Haemophilus influenzae hmw1 gene cluster.
KW   Non-typeable Haemophilus; high molecular weight surface protein;
KW   HMW1; hmw1 gene; immunogen; vaccine; otitis media; ss.
OS   Haemophilus influenzae strain 12.
FH   Key             Location/Qualifiers
FT   CDS             351. .4958
FT                   /*tag=  a
FT                   /note= "hmw1 gene"
FT   CDS             5114. .6748
FT                   /*tag=  b
FT                   /note= "ORF-b"
FT   CDS             7062. .9011
FT                   /*tag=  c
FT                   /note= "ORF-c"
PN   WO9736914-A1.
PD   09-OCT-1997.
PF   01-APR-1997; U04707.
PR   01-APR-1996; US-617697.
PA   (BARE/) BARENKAMP S J.
PI   Barenkamp SJ;
DR   WPI; 97-503038/46.
DR   P-PSDB; W30293.
PT   High molecular weight proteins of non-typeable Haemophilus
PT   influenzae - useful for vaccine production
PS   Disclosure; Page 78-82; 183pp; English.
CC   This nucleic acid comprises a gene cluster for the hmw1 gene (see
CC   also T90994) that encodes high molecular weight surface protein
CC   HMW1 (see W30293) of non-typeable Haemophilus influenzae strain 12.
CC   In addition to the hmw1 gene there are 2 additional downstream
CC   open reading frames that are required for the correct processing
CC   and secretion of the hmw1 gene product.  The ORF-b derived amino
CC   acid sequence demonstrates similarity with the derived amino acid
CC   sequences of 2 genes which encode proteins required for secretion
CC   and activation of haemolysins of P. mirabilis and S. marcescens.
CC   HMW proteins (see W30291-94) can be used in vaccines, as immunogens
CC   for preparation of antibodies and as antigens for detection of
CC   these antibodies.
SQ   Sequence   9171 BP;    3085 A;    1855 C;    1794 G;    2437 T;


alignment_scores:
           Quality:  102.50            Length:    353
             Ratio:  0.483              Gaps:     15
 Percent Similarity:  60.057   Percent Identity: 19.547

alignment_block:
 US-09-068-804-6 x T90996   ..

 Align seg 1/1  to: T90996  from: 1  to: 9171

      15 AsnAsnHisSerValGluAsnSerSerGlnThrAlaSerGlnSerValSe 31
         ||||||::::::   :::::::|||:::::::::|||:::   ::::::::
    3616 AATAACAACACTGAAGATAGCAGTGACAATAATGCCGGCTTAACTATCGA 3665

      31 rAlaLysAspIleLeuAsnSerIleGlyIleSerSerSerLysValSerA 48
         :||||||::::::      :::   :::|||:::|||   |||:::   :
    3666 TGCAAAAAATGTAACAGTAAACAACAATATTACTTCTCACAAAGCAGTGA 3715

      48 spLeuGly..................................... 50
         ::::::::
    3716 GCATCTCTGCGACAAGTGGAGAAATTACCACTAAAACAGGTACAACCATT 3765
```

```
      51 .....................................Le 51
                                             ||
    3766 AACGCAACCACTGGTAACGTGGAGATAACCGCTCAAACAGGTAGTATCCT 3815

      51 uSerProThrLeuSerAlaProAlaProGlyValLeuThrGlnThrProG 68
         |:::   :::   |||:::::::::::   :::||||||   |||   |
    3816 AGGTGGAATTGAGTCCAGCTCTGGCTCTGTAACACTTACTGCAACCGAGG 3865

      68 ly......ThrIleThrSerPheLeuLysAlaSerIleGlnAsnThrAsp 82
         ||      ::::::::::::::      :::::::::      |||:::
    3866 GCGCTCTTGCTGTAAGCAATATTTCGGGCAACACCGTTACTGTTACTGCA 3915

      83 MetAsnGlnAspLeuAsnAlaLeuAlaAsnAsnValThrThrLysAlaAs 99
         :::   :::|||::::::||||||::::::   |||:::|||:::::
    3916 AATAGCGGTGCATTAACCACTTTGGCAGGCTCT...ACAATTAAAGGAAC 3962

      99 nGluValValGlnThrGlnLeuArgGluGlnGlnAlaGluValGlyLysP 116
         :|||   |||   |||            |||   ::::::::::|||
    3963 CGAGAGTGTAACCACT......TCAAGTCAATCAGGCGATATCGGC.... 4002

     116 hePheAspIleSerGlyMetSerSerSerAlaValAlaLeuLeuAlaAla 132
                 :::   :::|||::::::::::|||::::::   |||:::
    4003 ..........GGTACGATTTCTGGTGGCACAGTAGAGGTTAAAGCAACC 4041

     133 AlaAsnThrLeuMetLeuThrLeuAsnGlnAlaAspSerLysLeuSerGl 149
         ::::::         ||||||   :::|||::::::||||||:::   ::
    4042 GAAAGT.........TTAACC...ACTCAATCCAATTCAAAAATTAAAGC 4079

     149 yLysLeuSerLeuValSerPheAspAlaAlaLysThrThrAlaSerSerM 166
         ::::   :::            :::|||:::   |||:::|||::::::
    4080 AACAACAGGC............GAGGCTAACGTAACAAGTGCAACAGGTA 4117

     166 etMetArgGluGlyMetAsnAlaLeuSerGlySerIleSerGlnSerAla 182
           :::         :::::::::   :::   :::::::::   :::|||
    4118 CAATTGGTGGTACGATTTCCGGTAATACGGTAAATGTTACGGCAAACGCT 4167

     183 LeuGlnLeuGlyIleThrGlyValGlyAlaLysLeuGluTyrLysGlyLe 199
         :::|||   :::   |||   ||||||   ::::::
    4168 GGCGATTTAACAGTT...GGGAATGGCGCAGAAATTAAT.........GC 4205

     199 uGlnAsnGluArgGlyAlaLeuLysHisAsnAlaAlaLysIleAspLysL 216
         :::      ::::::|||:::   :::::::::          ||||
    4206 GACAGAAGGAGCTGCAACCTTAACTACATCATCGGGC.........AAAT 4246

     216 euThrThrGluSerHisSerIleLysAsnValLeuAsnGlyGlnAsnSer 232
         ||||||||||||:::   |||      ::::::   ::::::::::::
    4247 TAACTACCGAAGCTAGTTCA......CACATTACTTCAGCCAAGGGTCAG 4290

     233 ValLysLeuGlyAlaGluGlyValAspSerLeuLysSerLeuAsnMetLy 249
         |||:::|||:::|||::::::   :::          |||:::|||
    4291 GTAAATCTTTCAGCTCAGGATGGTAGCGTTGCAGGAAGTATTAATGCCGC 4340

     249 sLysThrGlyThrAspAlaThrLysAsnLeuAsnAspAlaThrLeuLysS 266
         ::::::      ::::::|||   :::|||       :::|||:::|||:
    4341 CAATGTGACACTAAATACTACAGGCACTTTA......ACTACCGTGAAGG 4384

     266 erAsnAlaGlyThrSerAlaThrGluSerLeuGlyIleLysAspSerAsn 282
         ::::::::   ::::::::::::   :::|||   |||:::::::   :::
    4385 GTTCAAACATTAATGCAACCAGCGGTACCTTGGTTATTAACGCAAAAGAC 4434

     283 LysGlnIleSerProGluHisGlnAlaIleLeuSerLysArgLeuGluSe 299
         ::::::         :::   |||   |||::::::::::
    4435 GCTGAGCTA.........AATGGCGCAGCATTGGGTAACCACACAGTGGT 4475

     299 rValGluSerAspIleArgLeuGluGlnAsnThrMetAspMetThrArgI 316
         :::::::::                ::::::::::::   |||
    4476 AAATGCAACCAACGCAAATGGCTCCGGCAGCGTAATCGCGACAACC.... 4521

     316 leAspAlaArgLysMetGlnMetThrGlyAspLeuIleMetLysAsnSer 332
         :::   ::::::::::::|||||||||||||||||      |||:::
    4522 .....TCAAGCAGAGTGAACATCACTGGGGATTTAATCACAATAAATGGA 4566

     333 ValThrVal 335
```

```
    4485 GCTGAGCTA.........AATGGCGCAGCATTGGGTAACCACACAGTGGT 4525

     299 rValGluSerAspIleArgLeuGluGlnAsnThrMetAspMetThrArgI 316
         :::::::::              ::::::::::   |||
    4526 AAATGCAACCAACGCAAATGGCTCCGGCAGCGTAATCGCGACAACC.... 4571

     316 leAspAlaArgLysMetGlnMetThrGlyAspLeuIleMetLysAsnSer 332
            :::  ::::::::::::|||||||||||||||     |||:::
    4572 .....TCAAGCAGAGTGAACATCACTGGGGATTTAATCACAATAAATGGA 4616

     333 ValThrVal 335
         :::::::::
    4617 TTAAATATC 4625

seq_name: N_Geneseq_36:V22834

seq_documentation_block:
ID   V22834 standard; DNA; 8930 BP.
AC   V22834;
DT   19-AUG-1998  (first entry)
DE   Haemophilus paragallinarum antigenic protein encoding DNA #1.
KW   Haemophilus paragallinarum; antigenic protein; HI antibody; diagnosis;
KW   vaccine; chicken infectious coryza; CIC; fowl; ds.
OS   Haemophilus paragallinarum.
FH   Key             Location/Qualifiers
FT   CDS             243. .6371
FT                   /*tag= a
FT                   /note= "antigenic protein"
FT   sig_peptide     243. .452
FT                   /*tag= b
FT   mat_peptide     453. .6368
FT                   /*tag= c
FT   CDS             8375. .8930
FT                   /*tag= d
FT                   /note= "antigenic protein; no stop codon given"
PN   WO9812331-A1.
PD   26-MAR-1998.
PF   12-SEP-1997; J03222.
PR   19-SEP-1996; JP-271408.
PA   (KAGA ) ZH KAGAKU & KESSEI RYOHO KENKYUSHO.
PA   (KAGA ) CHEMO-SERO-THERAPEUTIC RES INST.
PI   Hamada F, Matsuo K, Sakaguchi M, Tokiyoshi S, Tokunaga E;
DR   WPI; 98-230318/20.
DR   P-PSDB; W56319, W56320.
PT   Antigenic polypeptide from Haemophilus paragallinarum induces HI
PT   antibody production - and is useful for diagnosis of and preparation
PT   of vaccines for chicken infectious coryza
PS   Claim 1; Page 51-69; 108pp; Japanese.
CC   The present sequence encodes two antigenic proteins derived from
CC   Haemophilus paragallinarum strain A-221. The antigenic proteins
CC   stimulate the production of HI antibodies in fowl. The proteins
CC   and DNA coding for them can be used in the preparation of vaccines
CC   for the prevention of chicken infectious coryza (CIC). The proteins
CC   and their antibodies can be used in the diagnosis and treatment of CIC.
SQ   Sequence   8930 BP;    2933 A;    1623 C;    1994 G;    2380 T;


alignment_scores:
            Quality:  102.50              Length:    344
              Ratio:  0.475                Gaps:     14
 Percent Similarity:  62.791   Percent Identity: 19.186

alignment_block:
 US-09-068-804-6 x V22834   ..

 Align seg 1/1  to: V22834  from: 1  to: 8930

      89 AlaLeuAlaAsnAsnValThrThrLysAlaAsnGluValValGlnThrGl 105
         :::::::::||||||:::::::::   :::|||:::      ::::::::
    5418 TCTATTGAGAATAACTTGGTAATTGGTTCAAATGATCCTGAGAAGGCAAA 5467

     105 nLeuArgGluGlnGlnAlaGluValGlyLysPhePheAspIleSerGlyM 122
         :|||   :::|||:::::::::::::         :::   |||:::::::

    5468 ATTAGCCGCACAAGAAGGTAATGCT.........TTGGTTATCACTAACA 5508

     122 etSerSerSerAlaValAlaLeuLeuAlaAlaAlaAsnThrLeuMetLeu 138
         :::::::::::::::::|||::::::   ::::::::::::   ||||||
    5509 AAGATGACGGGAATGCGGCGATGGTCTTTAATAACGAGAAAAATATGCTT 5558

     139 ThrLeuAsnGlnAlaAspSerLys.....................LeuSe 148
         :::|||::::::   ::::::|||                      ::
    5559 GTTCTCAGTGATAAAGAGGCGAAACCAAGAGTGCTTCTTGATGGACAAAA 5608

     148 rGlyLysLeuSerLeuValSerPheAspAlaAlaLysThrThrAlaSerS 165
         :|||   |||:::||||||:::   |||:::::::::::::|||   ||||
    5609 TGGGGCATTAACTTTAGTCGGCAATGATGATTCTCAAGTCACCCTTTCCT 5658

     165 erMetMetArgGluGlyMetAsnAlaLeuSerGlySerIleSerGlnSer 181
         ||      :::   |||::::::      :::::::::::::|||:::
    5659 CT......AAGAAAGGTAAAGATATTGATGGAAATGATTTGAGCCGTCTC 5702

     182 AlaLeuGlnLeuGlyIleThrGlyValGlyAlaLysLeuGlu........ 195
         ::::::           |||:::::::::::::::::||||||
    5703 TCTGTGACGACTGAAAGAACAAATGCTGATGGGCAACTTGAAAAAGTGGA 5752

     196 .......TyrLysGlyLeuGlnAsnGluArgGlyAlaLeuLysHisAsnA 210
             :::      :::::::::         :::||||||   :::|
    5753 AACCTCATTTGCTACAATGGATGAT.........GGCTTGAAGTTCAAAG 5793

     210 laAlaLys.........IleAspLysLeuThrThrGluSer......... 220
         ||:::            |||:::|||      :::|||:::
    5794 CCGACGGGGATAAAGTGATTAATAAGAAACTTAATGAAACCGTTGAAATT 5843

     221 .........HisSerIleLysAsnValLeuAsnGlyGlnAsnSerValLy 234
                 :::::::::::::::   :::::::::::|||   |||||
    5844 GTTGGTGATGAGAATGTGACAACATCTATTACTGATGATAATAAGGTGAA 5893

     234 sLeuGlyAlaGluGlyValAspSerLeuLysSerLeuAsnMetLysLysT 251
         |::::::   :::         ::::::      :::::::::   :::|
    5894 AGTTTCACTGAATAAGAAAATCGCGATTGATGAGGTTAAGATTCCAAATA 5943

     251 hrGlyThrAspAlaThrLysAsnLeuAsnAspAlaThrLeuLysSerAsn 267
         ||::::::||||||   |||      :::|||:::::::::   :::|||
    5944 CAGATCCTGATGCTCAAAAG......GGAGATAGCATTGTAATCAACAAT 5987

     268 AlaGlyThrSerAlaThrGluSerLeu......GlyIleLysAspSerAs 282
         :::|||:::   |||   :::   :::      |||:::|||:::|||::
    5988 GGTGGAATCCACGCAGGTAATAAAGTGATTACTGGCGTTAAAGCGAGTGA 6037

     282 nLysGlnIleSerProGluHisGlnAlaIleLeuSerLysArgLeuGluS 299
         :   :::::::|||:::   :::::::::   |||::::::   :::::::
    6038 TGACCCAACCAGTGCAGTGAATCGAGGTCAATTAAATACTGTGATTGATA 6087

     299 erValGluSerAspIle...ArgLeuGluGlnAsnThrMetAspMetThr 314
         ::|||::::::::::::   ::::::::::|||   :::   |||:::|||
    6088 ATGTTCAAAATAATTTCAATCAAGTTAATCAACGTATTGGCGATTTAACA 6137

     315 ArgIleAspAlaArgLysMetGlnMetThrGlyAspLeuIleMetLysAs 331
         |||   ::::::|||
    6138 CGG...GAGTCGCGT................................... 6149

     331 nSerValThrValGlyGlyIleAlaGlyAlaSerGlyGlnTyrAlaAlaT 348
          :::   ::::::|||:::::::|||   ||||||   |||   :::|||
    6150 .GCAGGTATTGCAGGTGCAATGGCGACGGCAAGCCTACAAAATGTTGCTT 6198

     348 hrGlnGluArgSerGluGlnGlnIleSerGlnValAsnAsnArgValAla 364
           :::   ::::::         ::::::   ::::::   :::   :::
    6199 TACCAGGGAAAACAACGATTTCCGTAGGTACAGCAACGTTCAAAGGGGAG 6248

     365 SerThrAla.....................SerAspGluAlaArgGluSe 374
         :::::::::                     ||||||::::::::::   ::
    6249 AATGCTGTTGCAATAGGGATGTCTAGACTCTCTGATAATGGAAAAGTAGG 6298

     374 rSerArgLysSerThrSerLeuIleGlnGluMetLeuLysThrMetGluS 391
         :   |||                        |||     |||   :
    6299 TATCCGT.........................TTATCTGGTATGAGTA 6321
```

```
  4297 TAACTACCGAAGCTAGTTCA......CACATTACTTCAGCCAAGGGTCAG 4340

   233 ValLysLeuGlyAlaGluGlyValAspSerLeuLysSerLeuAsnMetLy 249
       |||:::|||:::|||:::::   :::         |||:::|||
  4341 GTAAATCTTTCAGCTCAGGATGGTAGCGTTGCAGGAAGTATTAATGCCGC 4390

   249 sLysThrGlyThrAspAlaThrLysAsnLeuAsnAspAlaThrLeuLysS 266
        ::::::     :::::|||   :::|||      :::|||:::|||:
  4391 CAATGTGACACTAAATACTACAGGCACTTTA......ACTACCGTGAAGG 4434

   266 erAsnAlaGlyThrSerAlaThrGluSerLeuGlyIleLysAspSerAsn 282
       ::::::::   ::::::::::::   :::|||   |||::::::   :::
  4435 GTTCAAACATTAATGCAACCAGCGGTACCTTGGTTATTAACGCAAAAGAC 4484

   283 LysGlnIleSerProGluHisGlnAlaIleLeuSerLysArgLeuGluSe 299
          ::::::         :::   |||   |||:::::::::
  4485 GCTGAGCTA.........AATGGCGCAGCATTGGGTAACCACACAGTGGT 4525

   299 rValGluSerAspIleArgLeuGluGlnAsnThrMetAspMetThrArgI 316
          :::::::::             ::::::::::   |||
  4526 AAATGCAACCAACGCAAATGGCTCCGGCAGCGTAATCGCGACAACC.... 4571

   316 leAspAlaArgLysMetGlnMetThrGlyAspLeuIleMetLysAsnSer 332
          :::   ::::::::::|||||||||||||||     |||:::
  4572 .....TCAAGCAGAGTGAACATCACTGGGGATTTAATCACAATAAATGGA 4616

   333 ValThrVal 335
       :::::::::
  4617 TTAAATATC 4625

seq_name: N_Geneseq_36:Q72293

seq_documentation_block:
ID   Q72293 standard; DNA; 5116 BP.
AC   Q72293;
DT   25-JUN-1995  (first entry)
DE   Haemophilus high molecular weight protein HMW1 DNA sequence.
KW   High molecular weight protein; HMW1; protective vaccine; otitis;
KW   sinusitis; bronchitis; Hib; ss.
OS   Haemophilus.
PN   WO9421290-A.
PD   29-SEP-1994.
PF   15-MAR-1994; U02550.
PR   16-MAR-1993; US-038682.
PA   (BARE/) BARENKAMP S J.
PA   (SGEM/) ST GEME J W.
PI   Barenkamp SJ,  St GEME JW;
DR   WPI; 94-316665/39.
DR   P-PSDB; R63505.
PT   New immunogenic high mol. wt. proteins of non typeable
PT   Haemophilus - useful in protective vaccines
PS   Claim 2; Page 28; 127pp; English.
CC   The HMW1 protein encoded by this sequence is useful in a vaccine to
CC   protect against disease caused by non-typeable Haemophilus which are
CC   not controlled by H. influenzae type b (Hib) vaccines.  The encoded
CC   protein can also be used as a carrier for protective Hib
CC   polysaccharide (in a conjugate vaccin against meningitis) or for
CC   other antigens, haptens, etc.  This DNA sequence is part of a larger
CC   sequence encoding a gene cluster for the hmw1 gene (9011 bp).
SQ   Sequence   5116 BP;    1766 A;    1008 C;    1037 G;    1305 T;


alignment_scores:
            Quality:  102.50              Length:    353
              Ratio:  0.483                Gaps:     15
 Percent Similarity:  60.057   Percent Identity: 19.547

alignment_block:
 US-09-068-804-6 x Q72293   ..

 Align seg 1/1  to: Q72293  from: 1  to: 5116

      15 AsnAsnHisSerValGluAsnSerSerGlnThrAlaSerGlnSerValSe 31
```

```
       ||||||::::::   ::::::|||:::::::::|||:::   ::::::::
  3666 AATAACAACACTGAAGATAGCAGTGACAATAATGCCGGCTTAACTATCGA 3715

    31 rAlaLysAspIleLeuAsnSerIleGlyIleSerSerSerLysValSerA 48
       :||||||::::::      :::   :::|||:::|||   |||:::   :
  3716 TGCAAAAAATGTAACAGTAAACAACAATATTACTTCTCACAAAGCAGTGA 3765

    48 spLeuGly..................................... 50
       ::::::::
  3766 GCATCTCTGCGACAAGTGGAGAAATTACCACTAAAACAGGTACAACCATT 3815

    51 .............................................Le 51
                                                     ||
  3816 AACGCAACCACTGGTAACGTGGAGATAACCGCTCAAACAGGTAGTATCCT 3865

    51 uSerProThrLeuSerAlaProAlaProGlyValLeuThrGlnThrProG 68
       |:::   :::   |||:::::::::::::   :::||||||   |||   |
  3866 AGGTGGAATTGAGTCCAGCTCTGGCTCTGTAACACTTACTGCAACCGAGG 3915

    68 ly......ThrIleThrSerPheLeuLysAlaSerIleGlnAsnThrAsp 82
       ||       ::::::::::::::       :::::::::       |||:::
  3916 GCGCTCTTGCTGTAAGCAATATTTCGGGCAACACCGTTACTGTTACTGCA 3965

    83 MetAsnGlnAspLeuAsnAlaLeuAlaAsnAsnValThrThrLysAlaAs 99
          :::   :::|||::::::|||||||::::::   |||:::|||:::::
  3966 AATAGCGGTGCATTAACCACTTTGGCAGGCTCT...ACAATTAAAGGAAC 4012

    99 nGluValValGlnThrGlnLeuArgGluGlnGlnAlaGluValGlyLysP 116
       :|||   |||   |||          |||   ::::::::::|||
  4013 CGAGAGTGTAACCACT......TCAAGTCAATCAGGCGATATCGGC.... 4052

   116 hePheAspIleSerGlyMetSerSerSerAlaValAlaLeuLeuAlaAla 132
                  :::   :::|||:::::::::|||::::::   |||:::
  4053 ...........GGTACGATTTCTGGTGGCACAGTAGAGGTTAAAGCAACC 4091

   133 AlaAsnThrLeuMetLeuThrLeuAsnGlnAlaAspSerLysLeuSerGl 149
       ::::::         ||||||   :::|||:::::::||||||:::   ::
  4092 GAAAGT.........TTAACC...ACTCAATCCAATTCAAAAATTAAAGC 4129

   149 yLysLeuSerLeuValSerPheAspAlaAlaLysThrThrAlaSerSerM 166
       ::::   :::              :::|||:::   |||:::|||::::::
  4130 AACAACAGGC.............GAGGCTAACGTAACAAGTGCAACAGGTA 4167

   166 etMetArgGluGlyMetAsnAlaLeuSerGlySerIleSerGlnSerAla 182
         :::         :::::::::   :::   :::::::::   :::|||
  4168 CAATTGGTGGTACGATTTCCGGTAATACGGTAAATGTTACGGCAAACGCT 4217

   183 LeuGlnLeuGlyIleThrGlyValGlyAlaLysLeuGluTyrLysGlyLe 199
         :::|||   :::   |||   ||||||   ::::::
  4218 GGCGATTTAACAGTT...GGGAATGGCGCAGAAATTAAT.........GC 4255

   199 uGlnAsnGluArgGlyAlaLeuLysHisAsnAlaAlaLysIleAspLysL 216
          :::       ::::::|||:::   :::::::::          ||||
  4256 GACAGAAGGAGCTGCAACCTTAACTACATCATCGGGC.........AAAT 4296

   216 euThrThrGluSerHisSerIleLysAsnValLeuAsnGlyGlnAsnSer 232
       ||||||||||||:::   |||      ::::::   ::::::::::::
  4297 TAACTACCGAAGCTAGTTCA......CACATTACTTCAGCCAAGGGTCAG 4340

   233 ValLysLeuGlyAlaGluGlyValAspSerLeuLysSerLeuAsnMetLy 249
       |||:::|||:::|||:::::   :::         |||:::|||
  4341 GTAAATCTTTCAGCTCAGGATGGTAGCGTTGCAGGAAGTATTAATGCCGC 4390

   249 sLysThrGlyThrAspAlaThrLysAsnLeuAsnAspAlaThrLeuLysS 266
        ::::::     :::::|||   :::|||      :::|||:::|||:
  4391 CAATGTGACACTAAATACTACAGGCACTTTA......ACTACCGTGAAGG 4434

   266 erAsnAlaGlyThrSerAlaThrGluSerLeuGlyIleLysAspSerAsn 282
       ::::::::   ::::::::::::   :::|||   |||::::::   :::
  4435 GTTCAAACATTAATGCAACCAGCGGTACCTTGGTTATTAACGCAAAAGAC 4484

   283 LysGlnIleSerProGluHisGlnAlaIleLeuSerLysArgLeuGluSe 299
          ::::::         :::   |||   |||:::::::::
```

```
     325 yAspLeuIleMetLysAsn..............................  331
             |||   ||||||
    1600 .......ATTGCGAAGAACTTCCCCAAGCTCTCGAAAGTCATCTCGTCTC 1642

     332 ................SerValThrValGlyGlyIleAlaGlyAlaSer  342
                 -        :::|||   ||||||   ::::::::::|||:::
    1643 TTACCAGTAAATGGGTCACGGTTGGGGTTGGGGTTGTAGTTGCGGCGCCT 1692

     343 Gly.......................GlnTyrAlaAlaThrGlnGluAr  351
         :::                       |||   ::::::   |||:::
    1693 GCTCTCGGTAAAGGGATTATGCAAATGCAGCTCTCGGAGATGCAACAAAA 1742

     351 gSerGluGln.....................................  354
           :::|||
    1743 CGTCGCTCAATTTCAGAAAGAAGTCGGAAAACTGCAGGCTGCGGCTGATA 1792

     355 ................GlnIleSerGlnValAsnAsnArgValAlaSer  365
                         |||:::   |||   ::::::::::::||||||
    1793 TGATTTCTATGTTCACTCAATTTTGGCAACAGGCAAGTAAAATTGCCTCA 1842

     366 ThrAlaSerAspGluAlaArgGluSerSerArgLysSerThrSerLeuIl  382
         :::   :::::::|||:::   |||   :::::::|||:::|||   |||
    1843 AAACAAACAGGCGAGTCTAATGAAATGACTCAAAAAGCTACCAAGCTGGG 1892

     382 eGlnGluMetLeuLysThrMetGluSerIleAsnGlnSerLysAlaSerA  399
           ::::::||||||:::   :::::::|||:::   :::   |||:::|
    1893 CGCTCAAATCCTTAAAGCGTATGCCGCAATCAGCGGAGCCATCGCTGGCG 1942

     399 la 399
           ||
    1943 CA 1944
```

seq_name: N_Geneseq_36:Q49506

seq_documentation_block:

```
ID   Q49506 standard; DNA; 5116 BP.
AC   Q49506;
DT   26-APR-1994  (first entry)
DE   Sequence encoding high molecular weight protein 1 (HMW1).
KW   HMW; high molecular weight protein; virus; vaccine; influenza;
KW   epitope; immunity; haemophilus influenzae; ss.
OS   Haemophilus influenzae.
FH   Key             Location/Qualifiers
FT   cds             351. .4961
FT                   /*tag= a
FT                   /product= High molecular weight protein 1
FT   misc_difference 360. .362
FT                   /*tag= b
FT                   /transl_except= CTA encodes Isoleucine.
FT   misc_difference 642. .644
FT                   /*tag= c
FT                   /transl_except= GAU encodes Alanine.
FT   misc_difference 1437. .1439
FT                   /*tag= d
FT                   /transl_except= AAG encodes Asparagine.
FT   misc_difference 1674. .1676
FT                   /*tag= e
FT                   /transl_except= CCG encodes Phenylalanine.
PN   WO9319090-A.
PD   30-SEP-1993.
PF   16-MAR-1993; U02166.
PR   16-MAR-1992; GB-005704.
PA   (BARE/) BARENKAMP S J.
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PI   Barenkamp SJ;
DR   WPI; 93-320683/40.
DR   P-PSDB; R41723.
PT   High molecular weight surface proteins - of non-typeable
PT   haemophilus which exhibit immunogenic properties
PS   Claim 3; Figure 1; 100pp; English.
CC   The isolation and purification of the high molecular weight protein
CC   enables the identification of the major protective epitopes of the
CC   protein by conventional epitope mapping.  These epitopes can then be
CC   synthesised using standard techniques and incorporated into fully
CC   synthetic or recombinant vaccines.
SQ   Sequence   5116 BP;    1766 A;    1008 C;    1037 G;    1305 T;
```

```
alignment_scores:
            Quality:  102.50              Length:    353
              Ratio:  0.483                 Gaps:     15
 Percent Similarity:  60.057   Percent Identity: 19.547

alignment_block:
 US-09-068-804-6 x Q49506   ..

 Align seg 1/1  to: Q49506  from: 1  to: 5116

      15 AsnAsnHisSerValGluAsnSerSerGlnThrAlaSerGlnSerValSe 31
         ||||||:::::   :::::::|||:::::::::|||:::   ::::::::
    3666 AATAACAACACTGAAGATAGCAGTGACAATAATGCCGGCTTAACTATCGA 3715

      31 rAlaLysAspIleLeuAsnSerIleGlyIleSerSerSerLysValSerA 48
         :||||||:::::      :::   :::|||:::|||   |||:::   :
    3716 TGCAAAAAATGTAACAGTAAACAACAATATTACTTCTCACAAAGCAGTGA 3765

      48 spLeuGly.......................................... 50
         ::::::::
    3766 GCATCTCTGCGACAAGTGGAGAAATTACCACTAAAACAGGTACAACCATT 3815

      51 .............................................Le 51
                                                       ||
    3816 AACGCAACCACTGGTAACGTGGAGATAACCGCTCAAACAGGTAGTATCCT 3865

      51 uSerProThrLeuSerAlaProAlaProGlyValLeuThrGlnThrProG 68
         |:::   :::   |||:::::::::::::   :::||||||   |||   |
    3866 AGGTGGAATTGAGTCCAGCTCTGGCTCTGTAACACTTACTGCAACCGAGG 3915

      68 ly.....ThrIleThrSerPheLeuLysAlaSerIleGlnAsnThrAsp 82
         ||      :::::::::::::::      :::::::::      |||:::
    3916 GCGCTCTTGCTGTAAGCAATATTTCGGGCAACACCGTTACTGTTACTGCA 3965

      83 MetAsnGlnAspLeuAsnAlaLeuAlaAsnAsnValThrThrLysAlaAs 99
         :::   :::|||:::::||||||::::::   |||:::|||:::::
    3966 AATAGCGGTGCATTAACCACTTTGGCAGGCTCT...ACAATTAAAGGAAC 4012

      99 nGluValValGlnThrGlnLeuArgGluGlnGlnAlaGluValGlyLysP 116
         :|||   |||   |||            |||   :::::::::|||
    4013 CGAGAGTGTAACCACT......TCAAGTCAATCAGGCGATATCGGC.... 4052

     116 hePheAspIleSerGlyMetSerSerSerAlaValAlaLeuLeuAlaAla 132
                 :::   :::|||:::::::::::|||::::::   |||:::
    4053 ...........GGTACGATTTCTGGTGGCACAGTAGAGGTTAAAGCAACC 4091

     133 AlaAsnThrLeuMetLeuThrLeuAsnGlnAlaAspSerLysLeuSerGl 149
         ::::::         ||||||   :::|||:::::|||||:::   ::
    4092 GAAAGT.........TTAACC...ACTCAATCCAATTCAAAAATTAAAGC 4129

     149 yLysLeuSerLeuValSerPheAspAlaAlaLysThrThrAlaSerSerM 166
         ::::   :::            :::|||:::   |||:::|||::::::
    4130 AACAACAGGC............GAGGCTAACGTAACAAGTGCAACAGGTA 4167

     166 etMetArgGluGlyMetAsnAlaLeuSerGlySerIleSerGlnSerAla 182
           :::      :::::::::   :::   :::::::::   :::|||
    4168 CAATTGGTGGTACGATTTCCGGTAATACGGTAAATGTTACGGCAAACGCT 4217

     183 LeuGlnLeuGlyIleThrGlyValGlyAlaLysLeuGluTyrLysGlyLe 199
         :::|||   :::   |||   ||||||   ::::::
    4218 GGCGATTTAACAGTT...GGGAATGGCGCAGAAATTAAT.........GC 4255

     199 uGlnAsnGluArgGlyAlaLeuLysHisAsnAlaAlaLysIleAspLysL 216
         :::      :::::::|||:::   :::::::::          ||||
    4256 GACAGAAGGAGCTGCAACCTTAACTACATCATCGGGC.........AAAT 4296

     216 euThrThrGluSerHisSerIleLysAsnValLeuAsnGlyGlnAsnSer 232
         ||||||||||||:::   |||      ::::::   ::::::::::::
```

```
seq_documentation_block:
ID   T14618 standard; DNA; 1947 BP.
AC   T14618;
DT   11-NOV-1996  (first entry)
DE   DHFR/C. pneumoniae antigen fusion protein (polypeptide B) DNA.
KW   Polypeptide antigen; polypeptide B; strain YK41; plasmid; probe;
KW   pCPN533T; primer; assay; detection; antibody; diagnosis;
KW   infection; fusion protein; dihydrofolate reductase; DHFR;
KW   Chlamydia pneumoniae; ds.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   cds             1. .1947
FT                   /*tag=  a
FT                   /note= "STOP codon absent"
PN   WO9609320-A1.
PD   28-MAR-1996.
PF   20-SEP-1995; J01896.
PR   20-SEP-1994; JP-224711.
PR   28-APR-1995; JP-106011.
PR   28-APR-1995; JP-106006.
PR   28-APR-1995; JP-106010.
PR   28-APR-1995; JP-106008.
PR   28-APR-1995; JP-106009.
PA   (HITB ) HITACHI CHEM CO LTD.
PI   Izutsu H,  Matsumoto A,  Obara K;
DR   WPI; 96-188399/19.
DR   P-PSDB; R94584.
PT   Recombinant Chlamydia pneumoniae antigen and antibodies to it -
PT   used for detection and assay of C. pneumoniae e.g. in clinical
PT   diagnosis
PS   Claim 25; Pages 97-103; 128pp; Japanese.
CC   The present sequence encodes the dihydrofolate reductase (DHFR)/
CC   C. pneumoniae polypeptide antigen fusion protein, polypeptide B.
CC   C. pneumoniae strain YK41 was cultured and genomic DNA extracted
CC   to prep. a lambda gt11 DNA library. The library was then screened
CC   with an anti-YK41 monoclonal antibody (MAb), which was prepd. by
CC   fusing spleen cells from a mouse infected with YK41 with myeloma
CC   P3/NS1/1-Ag4-1 to produce a MAb expressing hydridoma. The DNA
CC   obtd. was then fused with DHFR DNA and the expression vector
CC   pADA431 to give pCPN533T. The plasmid was used to transform an
CC   E. coli host, which was cultured to give the antigenic polypeptide
CC   fusion protein, polypeptide B. Polypeptide B and primers and
CC   probes derived from its DNA can be used in assays for the
CC   detection of antigenic polypeptide antibodies and DNA,
CC   respectively, useful in the diagnosis of C. pneumoniae infection.
SQ   Sequence   1947 BP;    591 A;    452 C;    483 G;    421 T;


alignment_scores:
           Quality:   99.00              Length:    484
             Ratio:  0.401                Gaps:     15
 Percent Similarity:  51.033   Percent Identity: 18.182

alignment_block:
 US-09-068-804-6 x T14618   ..

 Align seg 1/1  to: T14618  from: 1  to: 1947

      2 LeuIleSerAsnValGlyIleAsnProAlaAlaTyrLeuAsnAsnHisSe 18
        :::   |||::::::   |||:::   |||:::      ::::::   ||
    637 ATGGAAAGCGATGCCACTATTGCTGGTGCTTCTGGAAAAGACAAAACTTC 686

     18 rValGluAsnSerSerGlnThrAlaSerGlnSerValSerAlaLysAspI 35
        |      :::   ::::::||||||:::|||         |||
    687 CTCGACTACAAAAACAGAAACAGCTCCACAACAGGGAGTTGCT....... 729

     35 leLeuAsnSerIleGlyIleSerSerSerLysValSerAspLeuGlyLeu 51
                   |||      ||||||                  |||
    730 ..........GCTGGGAAAGAATCCTCAGAAAGTCAAAAGGCAGGTGCT 768

     52 SerProThrLeuSerAlaProAlaProGlyValLeuThrGlnThrProGl 68
        ::::::   :::|||::::::|||:::   :::   :::::::|||:::
    769 GATACTGGAGTATCAGGAGCGGCTGCTACTACAGCATCAAATACTGCAAC 818

     68 yThrIleThrSerPheLeuLysAlaSerIleGlnAsnThrAspMetAsn. 84
          :::|||:::      :::::::::||||||:::::::::::::::::
    819 AAAAATTGCT......ATGCAGACCTCTATTGAAGAGGCGAGCAAAAGTA 862

     85 .............GlnAspLeuAsnAlaLeuAlaAsnAsnValThrThr 96
                    |||:::|||::::::|||:::::::::      :::
    863 TGGAGTCTACCTTACAGTCACTTCAAAGCCTCAGTGCCGCGCAAATGAAA 912

     97 LysAlaAsnGluValValGlnThrGlnLeuArgGluGlnGlnAlaGluVa 113
          :::::::::||||||   :::   |||       :::   :::
    913 GAAGTCGAAGCGGTTGTTGTTGCTGCCCTCTCAGGGAAAAGTTCGGGTTC 962

    113 lGlyLysPhePheAspIleSerGlyMetSerSerSerAlaValAlaLeuL 130
          :::|||   :::::::::::::::::::   ::::::|||:::
    963 CGCAAAA...TTGGAAACACCTGACCTCCCCAAGCCCGGGGTCACACCAA 1009

    130 euAlaAlaAlaAsnThrLeuMetLeuThrLeuAsnGlnAlaAspSerLys 146
          :::::::::      :::   |||:::|||:::::::|||      :::
   1010 GATCAGAGGTTATCGAAATCGGACTCGCGCTTGCTAAAGCAATTCAGACA 1059

    147 Leu......SerGlyLysLeuSerLeuValSerPheAspAlaAlaLysTh 161
        |||       :::   |||   :::|||   :::::::::::::::::::
   1060 TTGGGAGAAGCCACAAAATCTGCCTTATCTAACTATGCAAGTACACAAGC 1109

    161 rThrAlaSerSerMetMetArgGluGlyMetAsnAlaLeuSerGlySerI 178
        :   |||:::
   1110 ACAAGCAGAC..................................... 1119

    178 leSerGlnSerAlaLeuGlnLeuGlyIleThrGlyValGlyAlaLysLeu 194

   1119 ..................................................... 1119

    195 GluTyrLysGlyLeuGlnAsnGluArgGlyAlaLeuLysHisAsnAlaAl 211
                  |||:::::::::   :::|||      :::|||
   1120 ..............CAAACAAATAAACTAGGTCTAGAAAAGCAAGCGAT 1154

    211 aLysIleAspLysLeuThrThrGluSerHisSerIleLys.......... 224
          ||||||||||||         |||   :::   :::|||
   1155 AAAAATCGATAAAGAACGAGAAGAATACCAAGAGATGAAGGCTGCCGAAC 1204

    225 ....................AsnValLeuAsnGlyGlnAsnSerValLys 234
                            ::::::::::::      |||:::|||:::
   1205 AGAAGTCTAAAGATCTCGAAGGAACAATGGATACTGTCAATACTGTGATG 1254

    235 LeuGlyAlaGluGlyValAspSerLeuLysSerLeu.............. 246
        :::::::::      :::   ::::::   |||:::
   1255 ATCGCGGTTTCTGTTGCCATTACAGTTATTTCTATTGTTGCTGCTATTTT 1304

    247 .......AsnMetLysLysThrGlyThrAspAlaThrLysAsnLeuAsnA 261
              :::         :::|||   :::|||      ::::::::::
   1305 TACATGCGGAGCTGGACTCGCTGGACTCGCTGCGGGAGCTGCTGTAGGTG 1354

    261 spAlaThrLeuLysSerAsnAlaGlyThrSerAlaThrGluSerLeu... 276
        ::|||:::      ::::::||||||::::::|||:::   ::::::
   1355 CAGCGGCAGCTGGAGGTGCAGCAGGAGCTGCTGCCGCAACCACGGTAGCA 1404

    277 .................GlyIleLysAspSerAsnLysGln......Il 285
                          ::::::   ::::::   ||||||      ||
   1405 ACACAAATTACAGTTCAAGCTGTTGTCCAAGCGGTGAAACAAGCTGTTAT 1454

    285 eSerProGluHisGlnAlaIleLeuSerLysArgLeuGluSerValGluS 302
        |::::::   :::||||||||||   :::          ::::::|||   |
   1455 CACAGCTGTCAGACAAGCGATCACCGCGGCTATAAAAGCGGCTGTCAAAT 1504

    302 erAspIleArgLeuGluGlnAsnThrMet..................... 311
        ||:::|||:::          :::|||:::
   1505 CTGGAATAAAAGCATTTATCAAAACTTTAGTCAAAGCGATTGCCAAAGCC 1554

    312 .........AspMetThrArgIleAspAlaArgLysMetGlnMetThrGl 325
                :::::::::::::::   |||:::      ||||||
   1555 ATTTCTAAAGGAATCTCTAAGGTTTTCGCTAAGGGAACTCAAATG..... 1599
```

```
PN   DE19515514-C1.
PD   12-SEP-1996.
PF   27-APR-1995; 015514.
PR   27-APR-1995; DE-015514.
PA   (PRIV-) PRIVATES INST IMMUNOLOGIE & MOLEKULARGEN.
PI   Renz M,  Seelig HP;
DR   WPI; 96-403153/41.
DR   P-PSDB; W02258.
PT   DNA encoding nucleolar-endosomal auto-antigen - useful for exact
PT   diagnosis of rheumatic disease, in gene therapy and for removal of
PT   specific auto-antibodies
PS   Claim 2; Fig 2; 15pp; German.
CC   Transformed cells can be cultured to produce the antigen p162, for use
CC   in exact (differential) diagnosis of rheumatic disease, i.e. they
CC   can detect, in immunoassays, Western blots, etc., rheumatism-
CC   specific auto-antibodies.  The antigen can be used therapeutically,
CC   in the removal of auto-antibodies from the circulation, or when
CC   coupled to a cytotoxin, the elimination of auto-antibody-
CC   producing lymphocytes.
SQ   Sequence   4779 BP;    1916 A;    801 C;    1007 G;    1055 T;


alignment_scores:
            Quality:  104.50              Length:    459
              Ratio:  0.482                Gaps:     18
 Percent Similarity:  47.277   Percent Identity: 18.736

alignment_block:
 US-09-068-804-6 x T58751   ..

 Align seg 1/1  to: T58751  from: 1  to: 4779

      24 GlnThrAlaSerGlnSerValSerAlaLys..................... 33
         |||::: ||||||::::::|||||
    1550 CAGGCTCTTCAGCAAAGCACCACGGCAAAACTTCGAGAAGCTCAGAATGA 1599

      34 .......AspIleLeuAsnSerIleGlyIleSerSerSerLysValSerA 48
               :::::::|||     ||||||     :::   |||:::   :
    1600 TTTGGAACAAGTTCTACGTCAAATTGGCGATAAGGACCAAAAGATCCAGA 1649

      48 spLeuGlyLeuSerProThrLeuSerAlaProAlaProGlyValLeuThr 64
         ::|||                                  :::|||
    1650 ACCTT................................GAAGCTTTATTA 1666

      65 GlnThrProGlyThrIleThrSerPheLeuLys.................. 75
         |||::::::        :::|||:::|||
    1667 CAGAAGAGTAAAGAAAATATTTCATTACTAGAAAAAGAAAGAGAAGATCT 1716

      76 ....AlaSerIleGlnAsnThrAsp.......................M 83
             |||   ||||||:::   :::                       :
    1717 TTATGCAAAAATTCAGGCTGGTGAAGGAGAGACTGCTGTTCTTAACCAGT 1766

      83 etAsnGlnAspLeuAsnAlaLeuAlaAsnAsnValThr............. 95
         ::::::::      ::::::|||   :::::::||||||
    1767 TACAAGAAAAAAACCATACACTACAGGAGCAAGTAACTCAACTAACAGAG 1816

      96 ............................ThrLysAlaAsnGluValValGl 103
                                     ::::::|||:::|||   :::::
    1817 AAGCTGAAGAATCAGTCAGAAAGTCATAAACAAGCCCAGGAGAATTTGCA 1866

     103 nThrGlnLeuArgGluGlnGlnAlaGluValGlyLysPhePheAspIleS 120
         :   |||::::::|||||||:::|||:::
    1867 TGACCAGGTACAAGAGCAGAAGGCACAT...................... 1894

     120 erGlyMetSerSerSerAlaValAlaLeuLeuAlaAlaAlaAsnThrLeu 136
                                    |||   ||||||:::
    1895 ............................CTTAGAGCTGCACAAGACCGT 1915

     137 MetLeuThrLeuAsnGlnAlaAspSerLysLeuSerGlyLysLeuSer.. 152
         :::|||:::|||:::   :::   :::   |||:::::::::|||:::
    1916 GTCCTTTCCCTAGAAACTAGTGTCAATGAATTAAATAGTCAATTAAATGA 1965

     152 ..................................................... 152

    1966 AAGCAAGGAGAAGGTCTCCCAGCTTGACATACAGATTAAAGCCAAAACCG 2015

     153 .....LeuValSerPheAspAlaAlaLysThrThrAlaSerSerMetMet 167
               |||::::||   :::|||||||||||||:::      :::   :::
    2016 AACTATTACTATCAGCAGAAGCAGCAAAAACTGCTCAAAGAGCTGATCTT 2065

     168 ArgGluGlyMetAsnAlaLeuSerGlySerIleSerGlnSerAlaLeuGl 184
         ::::::   :::::::::      :::::::::   :::         ::
    2066 CAGAATCATTTGGACACAGCTCAAAATGCATTACAAGATAAACAGCAGGA 2115

     184 nLeuGly..............IleThrGlyValGlyAlaLysLeuGlu. 195
         :|||:::             :::      |||   |||||||||:::
    2116 GTTAAATAAGATTACTACTCAGTTGGATCAGGTCACTGCAAAGTTACAAG 2165

     195 ..................................................... 195

    2166 ACAAGCAAGAACATTGCAGTCAGCTGGAAAGTCATCTTAAAGAATATAAA 2215

     196 ......TyrLysGlyLeuGlnAsnGluArgGlyAlaLeuLysHisAsnAl 210
               |||   :::|||::::::         :::|||         ::
    2216 GAGAAATACCTCTCTTTAGAACAGAAAACCGAAGAGCTA.........GA 2256

     210 aAlaLysIleAspLysLeuThrThrGluSerHisSerIleLys....... 224
         ::::::::|||   ||||||   ::::::|||       :::|||
    2257 AGGTCAAATTAAGAAACTAGAAGCTGATAGTCTTGAAGTTAAAGCAAGCA 2306

     225 .....AsnValLeuAsnGly..............GlnAsnSerValLys 234
               ::::::|||::::::               |||   ::::::
    2307 AGGAGCAGGCTTTGCAAGATCTACAACAGCAAAGACAGCTGAACACAGAT 2356

     235 LeuGlyAlaGluGlyValAspSerLeuLysSerLeuAsnMetLysLysTh 251
         |||         :::::::::       |||   |||::::|||   |||
    2357 TTAGAGCTCAGAGCCACAGAATTGAGTAAACAACTTGAAATGGAGAAGGA 2406

     251 rGlyThrAspAlaThrLysAsnLeuAsnAspAlaThrLeuLysSerAsnA 268
         :::::::::|||:::        |||         ||||||:::|
    2407 AATAGTATCCAGTACAAGA......TTGGATCTACAGAAAAAATCTGAAG 2450

     268 laGlyThrSerAlaThrGluSerLeuGlyIleLysAspSerAsnLysGln 284
         ||      |||   ::::::   |||      ::::::   :::||||||
    2451 CCCTTGAAAGTATCAAGCAAAAGCTTACCAAGCAAGAGGAAGAAAAACAA 2500

     285 IleSerProGluHisGlnAlaIleLeuSerLysArgLeuGluSerValGl 301
                              ||||||   :::   :::|||::::::
    2501 .....................ATCCTGAAACAAGATTTTGAAACTTTAAG 2529

     301 uSerAspIleArgLeuGluGlnAsnThrMetAspMetThrArgIleAspA 318
         :::::::::::::::::::::::   ::::::   :::|||||||:::
    2530 TCAAGAAACAAAGATTCAGCATGAGGAATTGAAT...AACAGAATTCAAA 2576

     318 la..............ArgLysMetGlnMetThrGlyAspLeuIleMet 329
         ::               :::|||:::::|||      :::   :::|||
    2577 CAACAGTAACAGAACTACAAAAAGTGAAAATGGAGAAAGAAGCTTTAATG 2626

     330 LysAsnSerValThrValGlyGlyIleAlaGlyAlaSerGlyGlnTyrAl 346
         ::::::      ||||||
    2627 ACAGAGCTTTCTACAGTA.............................. 2644

     346 aAlaThrGlnGluArgSerGluGlnGlnIleSerGlnValAsnAsnArgV 363
                               :::::::::|||:::|||::::::   :
    2645 ...................AAGGACAAACTATCAAAAGTTTCTGATTCTT 2675

     363 alAlaSerThrAlaSerAspGluAlaArgGluSerSerArgLysSerThr 379
         ::   ::::::   |||:::   :::::::|||   ::::::||::::::
    2676 TGAAAAACTCTAAAAGTGAATTTGAAAAGGAG...AATCAGAAAGGAAAA 2722

     380 SerLeuIleGlnGluMetLeuLysThr 388
         :::   |||   ::::::   ||||||
    2723 GCCGCTATATTAGACTTGGAAAAAACT 2749

seq_name: N_Geneseq_36:T14618
```

```
KW   vaccine; otitis media; sinusitis; lower respiratory tract infection;
KW   immunity enhancer; immunoassay reagent; ds.
OS   Moraxella catarrhalis.
FH   Key             Location/Qualifiers
FT   CDS             600. .2432
FT                   /*tag= a
FT                   /product= "UspA2 antigen of strain TTA24"
PN   WO9828333-A2.
PD   02-JUL-1998.
PF   19-DEC-1997; U23930.
PR   20-DEC-1996; US-033598.
PA   (TEXA ) UNIV TEXAS SYSTEM.
PI   Aebi C, Cope LD, Fiske MJ, Fredenburg R, Hansen EJ,
PI   Maciver I;
DR   WPI; 98-377595/32.
DR   P-PSDB; W68206.
PT   New peptide(s) containing the core epitope of Moraxella catarrhalis
PT   Usp proteins - useful in, e.g. vaccines to prevent or treat M.
PT   catarrhalis infection, and antibodies for passive immunisation
PS   Claim 33; Pages 160-161; 237pp; English.
CC   This DNA encodes a UspA2 antigen of Moraxella catarrhalis strain
CC   TTA24. Nucleic acid sequences encoding the UspA1 and A2 antigens of
CC   M. catarrhalis isolates 035E, 046E, TTA24 and TTA37 can be used in
CC   genetic vaccination. An antigenic composition or vaccine containing
CC   antigenic peptides from UspA1 or UspA2 antigens are used to induce an
CC   immune response in mammals against M. catarrhalis and can be used to
CC   treat infections such as otitis media, sinusitis, lower respiratory
CC   tract infections. They can also be used as immunity enhancers for other
CC   bacterial, parasitic or viral antigens, to raise antibodies and as
CC   immunoassay reagents for detecting specific antibodies. The antibodies
CC   are useful for passive immunisation and as immunoassay reagents.
CC   Detection of the epitopic core sequence, by immunoassay or by PCR, is
CC   used to diagnose infection. The Usp antigens encoding nucleic acid
CC   sequences are also used to produce recombinant proteins and for screening
CC   for potential anti-M. catarrhalis agents, while their fragments are
CC   useful as diagnostic probes or primers or to isolate variant sequences.
SQ   Sequence   2673 BP;    913 A;    506 C;    558 G;    696 T;


alignment_scores:
           Quality:  102.00            Length:    336
             Ratio:  0.510               Gaps:     13
 Percent Similarity:  59.524   Percent Identity: 19.048

alignment_block:
 US-09-068-804-6 x V41346   ..

 Align seg 1/1  to: V41346  from: 1  to: 2673

      79 GlnAsnThrAspMetAsnGlnAspLeuAsnAlaLeuAlaAsnAsnValTh 95
         |||::::::|||:::::::::::|||::::::||||||::::::||||||
    1374 CAAAAAGCGGATCTTACAAAAGACATCAAAGCACTTGAAAGCAATGTCGA 1423

      95 rThrLysAlaAsnGluValValGlnThrGlnLeuArgGluGlnGlnAlaG 112
                    :::   :::      :::|||   :::|||:::|||:
    1424 AGAAGGTTTGTTGGAT...CTAAGCGGTCGTCTGCTTGATCAAAAAGCGG 1470

      112 luValGlyLysPhePheAspIleSerGlyMet...SerSerSerAlaVal 127
          :::::   |||      ||||||   ::::::   |||:::   :::
    1471 ATCTTACAAAA......GACATCAAAGCACTTGAAAGCAATGTCGAAGAA 1514

      128 AlaLeuLeuAlaAlaAlaAsnThrLeuMetLeuThrLeuAsnGlnAlaAs 144
          :::||||||:::   ::::::   |||:::        ::::::   ::
    1515 GGTTTGTTGGATCTAAGCGGTCGTCTGCTTGATCAAAAAGCGGATATTGC 1564

      144 pSerLysLeuSerGlyLysLeuSerLeuValSerPheAspAlaAlaLysT 161
          :   :::   ::::::      :::|||:::::::::::::::   :::
    1565 TCAAAACCAAACAGACATCCAAGATCTGGCCGCTTACAACGAGCTACAAG 1614

      161 hrThrAlaSerSerMetMetArgGluGlyMetAsnAlaLeuSerGlySer 177
                :::   :::      |||:::::::::||||||:::   :::
    1615 ACCAGTATGCTCAAAAGCAAACCGAAGCGATTGACGCTCTAAATAAAGCA 1664

      178 IleSerGlnSerAlaLeuGlnLeuGlyIleThrGlyValGlyAlaLysLe 194
          |||::::::::::         :::|||   ::::::::::|||
    1665 AGCTCTGAGAATACA......CAAAACATCGAAGATCTGGCCGCT..... 1703

      194 uGluTyrLysGlyLeuGlnAsnGluArgGlyAlaLeuLysHisAsnAlaA 211
              |||:::   |||||||::::::   :::      |||::::::::::
    1704 ....TACAATGAGCTACAAGATGCCTATGCC......AAACAGCAAACCG 1743

      211 laLysIleAspLysLeu...ThrThrGluSerHisSerIleLysAsnVal 226
          ::   ||||||   |||   ::::::   |||:::::::::::|||:::
    1744 AAGCGATTGACGCTCTAAATAAAGCAAGCTCTGAGAATACACAAAACATT 1793

      227 LeuAsnGlyGlnAsnSerValLysLeuGlyAlaGluGlyValAspSerLe 243
          :::::::|||:::::::::         |||:::::::::::::::::
    1794 GCTAAAAACCAAGCGGATATT.........GCTAATAACATCAACAATAT 1834

      243 uLysSerLeuAsnMetLysLysThrGly...ThrAspAlaThrLysAsnL 259
          :      |||:::   ::::::         :::::::::::|||:::|
    1835 CTATGAGCTGGCACAACAGCAAGATCAGCATAGCTCTGATATCAAAACCT 1884

      259 euAsnAspAlaThrLeuLysSerAsnAlaGlyThrSerAlaThrGluSer 275
          ||:::   |||:::      :::|||::::::      ||||||:::
    1885 TGGCAAAAGCAAGT...GCTGCCAATACTAATCGTATTGCTACTGCTGAA 1931

      276 LeuGlyIleLysAspSerAsnLysGlnIleSerProGluHisGlnAlaIl 292
          |||||||||   :::::::::|||
    1932 TTGGGCATCGCTGAGAACAAAAAA..................... 1955

      292 eLeuSerLysArgLeuGluSerValGluSerAspIleArgLeuGluGlnA 309
                                  :::::::::|||      :::|||:
    1956 .....GACGCTCAGATCGCCAAAGCACAAG 1980

      309 snThrMetAspMetThrArgIleAspAlaArg..... 319
          :::::   :::::::|||   ||||||:::
    1981 CGAATGCCAACAAAACTGCGATTGATGAAAACAAAGCATCTGCGGATACC 2030

      320 LysMetGlnMetThrGlyAspLeuIleMetLysAsnSerValThrValGl 336
          |||         |||:::|||   |||   ||||||:::   ::::::
    2031 AAGTTTGCAGCAACAGCAGACGCCATTACCAAAAATGGAAATGCTATCAC 2080

      336 yGlyIleAlaGlyAlaSerGlyGlnTyrAlaAlaThrGlnGluArgSerG 353
          |||   :::      :::   :::::::::   :::      :
    2081 TAAAAACGCAAAATCTATCACTGATTTGGGCACTAAAGTGGATGGTTTTG 2130

      353 luGlnGlnIleSerGlnValAsnAsnArgValAlaSerThrAlaSerAsp 369
          ::   :::::::::   ::::::::::::|||:::::::
    2131 ACGGTCGTGTAACTGCATTAGACACCAAAGTCAATGCC..... 2168

      370 GluAlaArgGluSerSerArgLysSerThrSerLeuIleGlnGluMetLe 386
                                                 ::::::      ::
    2169 .....TTTGATGGTCGTAT 2182

      386 uLysThrMetGluSerIleAsnGlnSerLysAlaSerAlaLeuAlaAlaI 403
          ::::::::::::::|||      ::::::      :::|||   ||||||:
    2183 CACAGCTTTAGACAGTAAAGTTGAAAACGGTATGGCTGCCCAAGCTGCCC 2232

      403 leAlaGly 405
          :::::|||
    2233 TAAGTGGT 2240

seq_name: N_Geneseq_36:T58751

seq_documentation_block:
ID   T58751 standard; cDNA; 4779 BP.
AC   T58751;
DT   09-MAR-1997  (first entry)
DE   Nucleolar/endosomal auto-antigen.
KW   Auto-antibody; p162; rheumatic disease; antigen; diagnosis;
KW   gene therapy; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             71. .4306
FT                   /*tag= a
```

```
PN   WO9822575-A2.
PD   28-MAY-1998.
PF   21-NOV-1997; U21347.
PR   14-OCT-1997; US-061953.
PR   22-NOV-1996; US-031626.
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (UYWI-) UNIV WISCONSIN.
PI   Choi GH, Dillon PJ, Welch RA;
DR   WPI; 98-312461/27.
PT   New isolated uropathogenic E. coli nucleotide sequences - used to
PT   develop products for the detection of pathogenic E. coli and to
PT   elicit an immune response to pathogenic E. coli
PS   Claim 21; Page 202-207; 250pp; English.
CC   This sequence represents a E. coli strain J96 contig containing
CC   pathogenicity island (PAI) sequences, and represents a nucleic acid
CC   molecule of the invention. PAIs are large fragments of DNA which comprise
CC   pathogenicity determinants. The sequences of the invention are taken from
CC   PAI IV and PAI V. PAI IV is located at approximately 64 min (near pheV)
CC   on the E. coli chromosome and is greater than 170 kb. PAI V is located at
CC   approximately 94 min (at pheR) on the E. coli chromosome and is
CC   approximately 160 kb in size. Antibodies specific to the proteins encoded
CC   by the PAI open reading frames of the invention can be used in kits to
CC   detect uropathogenic E. coli. The proteins are used in vaccines to elicit
CC   a protective immune response in an animal to the uropathogenic E. coli
CC   strain J96.
SQ   Sequence   9319 BP;    1843 A;    2764 C;    2360 G;    2347 T;


alignment_scores:
            Quality:  109.00              Length:    413
              Ratio:  0.480                Gaps:     19
 Percent Similarity:  54.964    Percent Identity: 20.581

alignment_block:
 US-09-068-804-6 x V31271/rev   ..

 Align seg 1/1  to reverse of: V31271  from: 1  to: 9319

      55 LeuSerAlaProAlaProGlyValLeuThrGlnThrProGlyThrIleTh 71
         |||:::|||:::|||:::|||   ||||||   |||   ::::::::::
    7925 CTTACTGCTTCGGCAACCGGTCAGTTGACCAGTACCGGCGATATCATGAG 7876

      71 r.........SerPheLeuLysAlaSerIleGlnAsnThr.......... 81
         :          :::   |||||||||:::::   ::::::
    7875 CCAGGGTGACACCACGCTGAAAGCAGCCACCATGGACAACCGGGGCAGTC 7826

      82 ..............AspMetAsnGlnAspLeuAsnAlaLeuAlaAsnAsn 93
                        ::::::   |||   |||:::|||:::|||:::
    7825 TGCTTTCGGCCGGCACGCTCTCCCTTGATGGAAACTCACTGGATAACAGC 7776

      94 ValThrThrLysAlaAsnGluValValGlnThrGlnLeuArgGluGlnGl 110
           |||:::::::::|||:::|||:::      |||      :::
    7775 GGCACTGTCCAGGGTAACCATGTCACGATTCGCCAG......AACAGTGT 7732

     110 nAlaGluValGlyLysPhePheAspIleSerGly.............. 121
          ::::::   |||:::::::   :::|||:::::
    7731 CACCAACAGTGGCACGCTCACCGGGATAGCCGCGCTGACGCTTGCCGCCC 7682

     122 ..MetSerSerSerAlaValAlaLeuLeuAlaAlaAlaAsnThrLeuMet 137
           |||:::|||:::      ||||||:::::::::::::::::::|||:::
    7681 GTATGGCATCCCCTCAACCTGCGCTGATGAATAACGGAGGTTCATTGCTG 7632

     138 LeuThrLeuAsnGlnAlaAspSerLysLeuSerGlyLysLeuSerLeuVa 154
                                    ||||||   |||::::::
    7631 ..........................ACCAGCGGCGATCTGACAATC.. 7611

     154 lSerPheAspAlaAlaLysThrThrAlaSerSerMetMetArgGluGlyM 171
               :::|||   :::::::::||||||      ::::::|||:
    7610 .........ACCGCAGGCAGTATTACCAGTTCCGGACACTGGCAGGGCA 7571

     171 etAsnAlaLeuSerGlySerIleSerGlnSerAlaLeuGlnLeuGlyIle 187
         ::   :::|||         |||:::   :::::|||
    7570 AACGGGTGCTG.........ATCACCGCAGACAGTCTG............ 7542

     188 ThrGlyValGlyAlaLysLeuGluTyrLysGlyLeuGlnAsnGluArgGl 204
                                                 ::::::   ||
    7541 .....................................GCAAACAGCGG 7531

     204 yAlaLeuLysHisAsnAlaAlaLysIleAspLysLeuThrThrGluSerH 221
         ||||::::::   :::::::::   :::::::::||||||   |||
    7530 GGCGATCCAGGCGGCTGACAGCCTGACTGCACGTCTGACGGGTGAG.... 7485

     221 isSerIleLysAsnValLeuAsnGlyGlnAsnSerValLysLeuGlyAla 237
                         |||   :::   ::::::   |||:::   :::
    7484 .................CTCGTCAGCACAGCAGGCAGCAAAGTCACCTCG 7452

     238 GluGlyValAspSerLeuLysSerLeuAsnMetLysLysThrGlyThrAs 254
         :::|||      :::|||   :::||||||:::   :::::::|||
    7451 AACGGTGAAATGGCGCTCAGTGCACTGAATTTAAGCAACAGCGGACAATG 7402

     254 pAlaThrLysAsnLeuAsnAspAlaThrLeuLysSerAsnAlaGlyThrS 271
           :::||||||||||         ||||||||:::|||:::   ||||
    7401 GATTGCAAAAAATCTG.........ACCCTGAAGGCGAACTCACTGACCA 7361

     271 erAlaThrGluSerLeuGlyIleLysAsp.........SerAsnLysGln 284
         |||||   :::      |||:::   :::          :::   :::|||
    7360 GTGCGGGTGACATCACCGGTGTGGATGCTCTCACGCTCACGGTGAATCAG 7311

     285 IleSerProGluHis.....................GlnAlaIleLeuSe 294
         :::      :::|||                          :::::::|::
    7310 ACGCTGCACAATCACGCGAGCGGAAAACTGCTCAGTGCAGGTGTGCTGAC 7261

     294 rLysArgLeuGluSerValGluSerAspIleArgLeuGluGlnAsnThrM 311
         :  :::   :::||||||   :::|||   :::|||:::   ||:::
    7260 GCTGAAGGCAGACAGTGTCACAAACGACGGGCAATTACAGGGAAATGCCA 7211

     311 etAspMetThr.......................ArgIleAspAlaArg 319
               :::|||                          ::::::::::::
    7210 CCACCATCACGGCAGGACAACTCACAAACGGCGGGCATCTGCAGGGCGAA 7161

     320 LysMetGlnMetThr............................GlyAs 326
         ::::::   ::::::                           |||
    7160 ACGCTGACGCTGGCCGCCTCCGGTGGCGTGAACAACCGTTCCGGTGGTGT 7111

     326 pLeuIleMetLysAsnSerValThrVal..................... 335
           |||:::   :::|||:::::::::|||
    7110 TCTGATGAGCCGGAATGCACTGAATGTCAGTACTGCGACCCTGAGTAACC 7061

     336 ..GlyGlyIleAlaGlyAlaSerGlyGlnTyrAlaAlaThrGlnGluArg 351
           |||   |||   |||::::::|||         ::::::   :::|||
    7060 AGGGCACGATACAGGGTGGTGGCGGGGTTTCCCTGAACGCCACTGACCGT 7011

     352 SerGluGlnGlnIleSerGlnValAsnAsnArgVal......AlaSerTh 366
                         |||   ::::::::::::       :::||||::
    7010 ..............CTGCAGAACGACGGCAAAATCCTCTCCGGCAGTAA 6976

     366 rAlaSerAspGluAlaArgGluSerSerArgLysSerThrSerLeuIleG 383
         :   :::      |||:::      :::   :::::::::::::|||:::|
    6975 CCTCACGCTGACGGCGCAGGTGCTGGCGAACACCGGCAGCGGACTGGTAC 6926

     383 lnGlu.........MetLeuLysThrMetGluSerIleAsnGlnSerLys 396
         ||:::         :::|||   ::::::::::::::::|||   ::::::
    6925 AGGCTGCCACCCTGCTGCTGGATGTGGTGAATACTGTCAACGGCGGACGC 6876

     397 AlaSerAlaLeuAlaAlaIleAlaGlyAsnIleArgAla 409
               :::||||||:::   :::::::::::::::::
    6875 ......GTACTTGCCACCGGCAGTGCCGACGTTAAAGGA 6843

seq_name: N_Geneseq_36:V41346

seq_documentation_block:
ID   V41346 standard; DNA; 2673 BP.
AC   V41346;
DT   07-OCT-1998  (first entry)
DE   M. catarrhalis strain TTA24 UspA2 antigen encoding DNA.
KW   Moraxella catarrhalis; UspA1; UspA2; antigen; genetic vaccination;
```

```
PF   21-NOV-1997; U21347.
PR   14-OCT-1997; US-061953.
PR   22-NOV-1996; US-031626.
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (UYWI-) UNIV WISCONSIN.
PI   Choi GH, Dillon PJ, Welch RA;
DR   WPI; 98-312461/27.
PT   New isolated uropathogenic E. coli nucleotide sequences - used to
PT   develop products for the detection of pathogenic E. coli and to
PT   elicit an immune response to pathogenic E. coli
PS   Claim 21; Page 103-115; 250pp; English.
CC   This sequence represents a E. coli strain J96 contig containing
CC   pathogenicity island (PAI) sequences, and represents a nucleic acid
CC   molecule of the invention. PAIs are large fragments of DNA which comprise
CC   pathogenicity determinants. The sequences of the invention are taken from
CC   PAI IV and PAI V. PAI IV is located at approximately 64 min (near pheV)
CC   on the E. coli chromosome and is greater than 170 kb. PAI V is located at
CC   approximately 94 min (at pheR) on the E. coli chromosome and is
CC   approximately 160 kb in size. Antibodies specific to the proteins encoded
CC   by the PAI open reading frames of the invention can be used in kits to
CC   detect uropathogenic E. coli. The proteins are used in vaccines to elicit
CC   a protective immune response in an animal to the uropathogenic E. coli
CC   strain J96.
SQ   Sequence   22671 BP;    5697 A;    5544 C;    6050 G;    5373 T;


alignment_scores:
            Quality:  118.00              Length:    403
              Ratio:  0.524                 Gaps:     18
 Percent Similarity:  55.831   Percent Identity: 20.596

alignment_block:
 US-09-068-804-6 x V31200   ..

 Align seg 1/1  to: V31200  from: 1  to: 22671

      55 LeuSerAlaProAlaProGlyValLeuThrGlnThrProGlyThrIleTh 71
         |||:::|||:::|||:::|||   ||||||   |||   ::::::::::
   15332 CTTACTGCTTCGGCAACCGGTCAGTTGACCAGTACCGGCGATATCATGAG 15381

      71 r.........SerPheLeuLysAlaSerIleGlnAsnThr.......... 81
         :         :::   ||||||||||::::::   ::::::
   15382 CCAGGGTGACACCACGCTGAAAGCAGCCACCACGGACAACCGGGGCAGTC 15431

      82 ..............AspMetAsnGlnAspLeuAsnAlaLeuAlaAsnAsn 93
                        ::::::   |||   |||:::|||:::|||:::
   15432 TGCTTTCGGCCGGCACGCTCTCCCTTGATGGAAACTCACTGGATAACAGC 15481

      94 ValThrThrLysAlaAsnGluValValGlnThrGlnLeuArgGluGlnGl 110
            |||:::::::::::::::|||:::      |||      :::
   15482 GGCACTGTCCAGGGTGACCATGTCACGATTCGCCAG......AACAGTGT 15525

     110 nAlaGluValGlyLysPhePheAspIleSerGly.............. 121
          ::::::   |||::::::   :::|||::::::
   15526 CACCAACAGTGGCACGCTCACCGGGATCGCCGCGCTGACGCTTGCCGCCC 15575

     122 ..MetSerSerSerAlaValAlaLeuLeuAlaAlaAlaAsnThrLeuMet 137
           |||   |||:::      ||||||:::::::::::::::::::|||:::
   15576 GTATGGTATCCCCTCAACCTGCGCTGATGAATAACGGAGGTTCATTGCTG 15625

     138 ............LeuThrLeuAsn........................ 141
                     ||||||::::::
   15626 ACCAGCGGCGATCTGACAATCACCGCAGGCAGTCTGGTAAACAGCGGGGC 15675

     142 ....GlnAlaAspSerLysLeuSerGlyLysLeuSer......LeuValS 155
             ||||||::::::   |||:::::::::|||:::      |||||||
   15676 GATCCAGGCGGCTGACAGCCTGACTGCACGTCTGACGGGTGAGCTCGTCA 15725

     155 erPheAspAlaAlaLysThrThrAlaSerSerMetMetArgGluGlyMet 171
         ||   :::::::::|||:::|||::::::::::   |||
   15726 GCACAGCGGGCAGCAAAGTCACCTCGAACGGTGAAATG............ 15763

     172 AsnAlaLeuSerGlySerIleSerGlnSerAlaLeuGlnLeuGlyIleTh 188
                                  :::   ||||||||:::|||:::   ::
   15764 .....................GCGCTCAGTGCACTGAATTTAAGCAACAG 15792

     188 rGly......ValGlyAlaLysLeuGluTyrLysGlyLeuGlnAsnGluA 203
         :|||      ::::::   :::|||      |||:::
   15793 CGGACAATGGATTGCAAAAAATCTGACCCTGAAGGCG.............. 15829

     203 rgGlyAlaLeuLysHisAsnAlaAlaLysIleAspLysLeuThrThrGlu 219
           ::::::|||:::   :::|||:::   |||       :::   |||
   15830 ..AACTCACTGACC...AGTGCGGGTGACATCACCGGTGTGGATACTCTC 15874

     220 SerHisSerIleLysAsnValLeuAsnGlyGlnAsnSerValLysLeuGl 236
         :::   ::::::::::::::::||||||:::|||::::::   ||||||
   15875 ACGCTCACGGTGAATCAGACGCTGAACAATCAGGCGAACGGAAAACTGCT 15924

     236 yAlaGluGlyValAspSerLeuLysSerLeuAsnMetLysLysThrGlyT 253
          ::::::||||||   :::||||||:::   :::::::::
   15925 CAGTGCAGGTGTGCTGACGCTGAAGGCAGACAGTGTCACA.......... 15964

     253 hrAspAlaThrLysAsnLeuAsnAspAlaThrLeuLysSerAsnAlaGly 269
                           |||||||:::   |||::::::|||||||
   15965 ...................AACGACGGGCAATTACAGGGAAATGCCACC 15994

     270 ThrSerAlaThrGluSerLeuGlyIleLysAspSerAsnLysGlnIleSe 286
         |||   ::::::      |||
   15995 ACCATCACGGCAGGACAACTC............................. 16015

     286 rProGluHisGlnAlaIleLeuSerLysArgLeuGluSerValGluSerA 303
                                                      ::::
   16016 ........................................ACAAACG 16022

     303 spIleArgLeuGluGlnAsnThrMetAspMetThrArgIleAspAlaArg 319
         ::   :::|||:::   :::|||:::   ::::::      ::::::
   16023 GCGGGCATCTGCAGGGCGAAACGCTGACGCTGGCCGCCTCCGGTGGCGTG 16072

     320 LysMetGlnMetThrGlyAspLeuIleMetLysAsnSerValThrVal.. 335
         :::   :::      |||   |||:::   :::|||::::::::|||
   16073 AACAACCGTTCCGGTGGTGTTCTGATGAGCCGGAATGCACTGAATGTCAG 16122

     336 .....................GlyGlyIleAlaGlyAlaSerGlyGlnT 345
                               |||   |||   |||:::::::|||
   16123 TACTGCGACCCTGAGTAACCAGGGCACGATACAGGGTGGTGGCGGGGTTT 16172

     345 yrAlaAlaThrGlnGluArgSerGluGlnGlnIleSerGlnValAsnAsn 361
           ::::::   :::|||                  |||   ::::::
   16173 CCCTGAACGCCACTGACCGT..............CTGCAGAACGACGGC 16207

     362 ArgVal......AlaSerThrAlaSerAspGluAlaArgGluSerSerAr 376
         ::::::      :::|||:::   :::      |||:::      :::
   16208 AAAATCCTCTCCGGCAGTAACCTCACGCTGACGGCGCAGGTGCTGGCGAA 16257

     376 gLysSerThrSerLeuIleGlnGlu.........MetLeuLysThrMetG 390
         :::::::::::::|||:::|||:::         :::|||   :::::::
   16258 CACCGGCAGCGGACTGGTACAGGCTGCCACCCTGCTGCTGGATGTGGTGA 16307

     390 luSerIleAsnGlnSerLysAlaSerAlaLeuAlaAlaIleAlaGlyAsn 406
         :::::::::|||   ::::::      :::|||||||:::   ::::::::
   16308 ATACTGTCAACGGCGGACGC......GTACTTGCCACCGGCAGTGCCGAC 16351

     407 IleArgAla 409
         :::::::::
   16352 GTTAAAGGA 16360

seq_name: N_Geneseq_36:V31271

seq_documentation_block:
ID   V31271 standard; DNA; 9319 BP.
AC   V31271;
DT   01-OCT-1998  (first entry)
DE   E. coli J96 pathogenicity island contig #85.
KW   PAI; pathogenicity island; uropathogenic E. coli detection; PAI IV; pheR;
KW   PAI V; pheV; vaccine; protective immune response; ds.
OS   Escherichia coli.
```

```
alignment_scores:
            Quality:  109.50              Length:    516
              Ratio:  0.429                Gaps:     22
 Percent Similarity:  49.419   Percent Identity: 19.574

alignment_block:
 US-09-068-804-6 x T42166   ..

 Align seg 1/1  to: T42166  from: 1  to: 3319

       2 LeuIleSerAsnValGlyIleAsnProAlaAlaTyrLeuAsnAsnHisSe 18
         :::|||:::::::|||:::   ::::::   :::|||
    1547 ATGATCGCGACTGTGAGCCCGAGCGCGCTGAACTAC.............. 1582

      18 rValGluAsnSerSerGlnThrAlaSerGlnSerValSerAlaLysAspI 35
             |||::::::      |||          :::      |||:::||||
    1583 ....GAGGAGACGCTGAGCACGCTGCGGTACGCGTCGCGCGCGCGACA 1628

      35 leLeuAsnSerIleGlyIleSer............SerSerLysValSer 47
         ||:::|||          ::::::          :::    ::::::
    1629 TTGTGAATGTTGCGCAGGTGAACGAGGACCCGCGCGCACGGCGGATCCGC 1678

      48 AspLeu..............GlyLeuSerProThrLeuSerAlaProAl 59
         :::|||               ::::::   ::::::::::::::::   ::
    1679 GAGCTGGAGGAGCAGATGGAGGACATGCGGCAGGCGATGGCTGGCGGCGA 1728

      59 aProGlyValLeuThrGlnThrProGlyThrIleThrSerPheLeuLysA 76
         :|||:::   :::::::::          ::::::   ::::::|||   :
    1729 CCCCGCGTACGTGTCTGAGCTGAAGAAGAAGCTT...GCGCTGCTGGAGT 1775

      76 laSerIleGlnAsnThrAspMetAsnGlnAspLeuAsnAlaLeuAlaAsn 92
         ::      |||:::          :::   ||||||:::||||||:::
    1776 CGGAGGCGCAGAAG......CGTGCGGCGGACCTGCAGGCGCTGGAGAGG 1819

      93 AsnValThrThrLysAlaAsnGluValValGlnThrGln...LeuArgGl 108
         :::          ::::::::::::   ||||||   :::   ||||||::
    1820 GAG.........CGGGAGCACAACCAGGTGCAGGAGCGGCTGCTGCGCGC 1860

     108 uGlnGlnAlaGluValGlyLysPhePheAspIleSerGlyMetSerSerS 125
         :   :::||||||                  |||   :::   |||
    1861 GACGGAGGCGGAG..................AAGAGCGAGCTGGAGTCGC 1892

     125 erAlaValAlaLeu.........LeuAlaAlaAla.............. 133
           |||:::||||||         ::::::|||:::
    1893 GTGCGGCTGCGCTGCAGGAGGAGATGACCGCGACTCGACGGCAGGCGGAC 1942

     134 ......AsnThrLeuMetLeuThrLeuAsnGlnAlaAspSerLysLeuSe 148
               :::::::|||   |||   |||::::::::::::::::::
    1943 AAGATGCAGGCGCTGAACCTGCGGCTGAAGGAAGAGCAGGCGCGCAAGGA 1992

     148 rGlyLysLeu......SerLeuValSerPheAspAlaAla.......... 159
               |||          ::::::      |||||||||
    1993 GCGCGAGCTGCTGAAAGAGATGGCGAAGAAGGACGCCGCGCTCTCGAAGG 2042

     160 ..............................................Lys 160

    2043 TTCGGCGACGCAAAGACGCCGAGATAGCAAGCGAGCGCGAGAAGCTGGAG 2092

     161 ThrThrAlaSerSerMetMetArgGlu..................GlyMe 171
         :::|||::::::   :::   ||||||                  :::::
    2093 TCGACCGTGGCGCAGCTGGAGCGTGAGCAGCGCGAGCGCGAGGTGGCTCT 2142

     171 tAsnAlaLeuSerGlySerIleSer......GlnSerAlaLeuGlnLeuG 186
         :::::|||||                    |||   ||||||:::   :
    2143 GGACGCATTGCAGACGCACCAGAGAAAGCTGCAGGAAGCGCTCGAGAGCT 2192

     186 lyIleThrGlyValGlyAlaLys............LeuGluTyrLysGly 198
         ::       ::::::::::::              :::   :::
    2193 CTGAGCGGACAGCCGCGGAAAGGGACCAGCTGCTGCAGCAGCTAACAGAG 2242

     199 LeuGlnAsnGluArgGlyAlaLeuLysHisAsnAlaAlaLysIleAspLy 215
         ||||||:::|||||||      |||   :::   ::::::      :::::
    2243 CTTCAGTCTGAGCGTACGCAGCTATCACAGGTTGTGACCGACCGCGAGCG 2292

     215 sLeuThrThrGluSerHisSerIleLys..................... 224
         :||||||   :::   :::   |||:::
    2293 GCTTACACGCGACTTGCAGCGTATTCAGTACGAGTACGGGGAAACCGAGC 2342

     224 ..............................................Lys 224

    2343 TCGCGCGAGACGTGGCGCTGTGCGCCGCGCAGGAGATGGAGGCGCGCTAC 2392

     225 ......AsnValLeuAsnGlyGlnAsnSerValLysLeuGlyAlaGluGl 239
               :::|||::::::   |||:::   :::   |||:::::|||
    2393 CACGCTGCTGTGTTTCACCTGCAAACGCTCCTGGAGCTCGCAACCGAGTG 2442

     239 yValAspSerLeuLysSerLeuAsnMetLysLysThrGlyThrAspAlaT 256
             |||:::|||:::       ::::::          :::   :::|||:
    2443 GGAGGACGCACTCCGCGAGCGTGCGCTTGCAGAGCGTGACGAAGCCGCTG 2492

     256 hrLysAsnLeuAsnAspAlaThrLeuLysSer...AsnAlaGlyThrSer 271
         ::   :::|||::::::|||:::   :::|||   ||||||       |||
    2493 CAGCCGAACTTGATGCCGCAGCCTCTACTTCCCAAAACGCACGTGAAAGC 2542

     272 AlaThrGluSerLeuGlyIleLysAspSerAsnLysGlnIleSerProGl 288
         |||   |||   |||          :::   :::   :::   |||   ||
    2543 GCCTGCGAGCGGCTAACCAGCCTTGAGCAGCAGCTTCGCGAATCCGAGGA 2592

     288 uHisGlnAlaIleLeuSerLysArgLeuGluSerValGluSerAspIleA 305
         |:::   |||   |||:::   :::|||||||:::::::::::::::
    2593 GCGCGCTGCGGAGCTGGCGAGCCAGCTGGAGGCCACTGCTGCTGCGAAGT 2642

     305 rgLeuGluGlnAsnThrMetAspMetThrArgIle.............. 316
             :::::::::       :::   ||||||
    2643 CGTCGGCGGAGCAGGACCGCGAGAACACGAGGGCCACGCTAGAGCAGCAG 2692

     317 ............AspAlaArgLysMetGlnMetThrGlyAspLeuIleMe 329
                       :::|||||||      ::::::::::::::|||
    2693 CTTCGCGAATCCGAGGCGCGCGCTGCGGAGCTGGCGAGCCAGCTG..... 2737

     329 tLysAsnSerValThrValGlyGlyIleAlaGlyAlaSerGlyGlnTyrA 346
                                       |||   |||::::::::::   :
    2738 .........................GAGGCCACTGCTGCTGCGAAGATGT 2762

     346 laAlaThrGlnGluArgSer.................GluGlnGlnIle 356
         ::|||   |||:::|||                   ||||||||||:::
    2763 CAGCGGAGCAGGACCGCGAGAACACGAGGGCCACGCTAGAGCAGCAGCTT 2812

     357 SerGlnValAsnAsnArgValAlaSerThrAlaSer.............. 368
            :::   :::::::|||:::|||      ||||||
    2813 CGTGACTCCGAGGAGCGCGCTGCGGAGCTGGCGAGCCAGCTGGAGTCCAC 2862

     369 ......................AspGluAlaArgGluSerSerArgLysS 378
                                :::::::::||||||||||:::|||
    2863 TACTGCTGCGAAGATGTCAGCGGAGCAGGACCGCGAGAGCACGAGG.... 2908

     378 erThrSerLeuIleGlnGluMetLeuLysThrMetGluSerIleAsnGln 394
           ::::::|||   |||::::::     :::   |||      ::::::
    2909 ..GCCACGCTAGAGCAGCAGCTTCGTGACTCCGAGGAGCGCGCTGCGGAG 2956

     395 SerLysAlaSerAlaLeuAlaAlaIleAlaGlyAsnIleArgAla*** 410
             :::         :::::::::|||:::::::::   ||||||
    2957 CTGGCGAGCCAGCTGGAGTCCACTACTGCTGCGAAGATGTCAGCGGAG 3004

seq_name: N_Geneseq_36:V31200

seq_documentation_block:
ID   V31200 standard; DNA; 22671 BP.
AC   V31200;
DT   01-OCT-1998  (first entry)
DE   E. coli J96 pathogenicity island contig #14.
KW   PAI; pathogenicity island; uropathogenic E. coli detection; PAI IV; pheR;
KW   PAI V; pheV; vaccine; protective immune response; ds.
OS   Escherichia coli.
PN   WO9822575-A2.
PD   28-MAY-1998.
```

```
  59 aProGlyValLeuThrGlnThrProGlyThrIleThrSerPheLeuLysA 76
1729 CCCCGCGTACGTGTCTGAGCTGAAGAAGAAGCTT...GCGCTGCTGGAGT 1775

  76 laSerIleGlnAsnThrAspMetAsnGlnAspLeuAsnAlaLeuAlaAsn 92
1776 CGGAGGCGCAGAAG......CGTGCGGCGGACCTGCAGGCGCTGGAGAGG 1819

  93 AsnValThrThrLysAlaAsnGluValValGlnThrGln...LeuArgGl 108
1820 GAG.........CGGGAGCACAACCAGGTGCAGGAGCGGCTGCTGCGCGC 1860

 108 uGlnGlnAlaGluValGlyLysPhePheAspIleSerGlyMetSerSerS 125
1861 GACGGAGGCGGAG..................AAGAGCGAGCTGGAGTCGC 1892

 125 erAlaValAlaLeu.........LeuAlaAlaAla.............. 133
1893 GTGCGGCTGCGCTGCAGGAGGAGATGACCGCGACTCGACGGCAGGCGGAC 1942

 134 ......AsnThrLeuMetLeuThrLeuAsnGlnAlaAspSerLysLeuSe 148
1943 AAGATGCAGGCGCTGAACCTGCGGCTGAAGGAAGAGCAGGCGCGCAAGGA 1992

 148 rGlyLysLeu......SerLeuValSerPheAspAlaAla.......... 159
1993 GCGCGAGCTGCTGAAAGAGATGGCGAAGAAGGACGCCGCGCTCTCGAAGG 2042

 160 .....Lys 160
2043 TTCGGCGACGCAAAGACGCCGAGATAGCAAGCGAGCGCGAGAAGCTGGAG 2092

 161 ThrThrAlaSerSerMetMetArgGlu.................GlyMe 171
2093 TCGACCGTGGCGCAGCTGGAGCGTGAGCAGCGCGAGCGCGAGGTGGCTCT 2142

 171 tAsnAlaLeuSerGlySerIleSer......GlnSerAlaLeuGlnLeuG 186
2143 GGACGCATTGCAGACGCACCAGAGAAAGCTGCAGGAAGCGCTCGAGAGCT 2192

 186 lyIleThrGlyValGlyAlaLys............LeuGluTyrLysGly 198
2193 CTGAGCGGACAGCCGCGGAAAGGGACCAGCTGCTGCAGCAGCTAACAGAG 2242

 199 LeuGlnAsnGluArgGlyAlaLeuLysHisAsnAlaAlaLysIleAspLy 215
2243 CTTCAGTCTGAGCGTACGCAGCTATCACAGGTTGTGACCGACCGCGAGCG 2292

 215 sLeuThrThrGluSerHisSerIleLys..................... 224
2293 GCTTACACGCGACTTGCAGCGTATTCAGTACGAGTACGGGGAAACCGAGC 2342

 224 ..... 224
2343 TCGCGCGAGACGTGGCGCTGTGCGCCGCGCAGGAGATGGAGGCGCGCTAC 2392

 225 ......AsnValLeuAsnGlyGlnAsnSerValLysLeuGlyAlaGluGl 239
2393 CACGCTGCTGTGTTTCACCTGCAAACGCTCCTGGAGCTCGCAACCGAGTG 2442

 239 yValAspSerLeuLysSerLeuAsnMetLysLysThrGlyThrAspAlaT 256
2443 GGAGGACGCACTCCGCGAGCGTGCGCTTGCAGAGCGTGACGAAGCCGCTG 2492

 256 hrLysAsnLeuAsnAspAlaThrLeuLysSer...AsnAlaGlyThrSer 271
2493 CAGCCGAACTTGATGCCGCAGCCTCTACTTCCCAAAACGCACGTGAAAGC 2542

 272 AlaThrGluSerLeuGlyIleLysAspSerAsnLysGlnIleSerProGl 288
2543 GCCTGCGAGCGGCTAACCAGCCTTGAGCAGCAGCTTCGCGAATCCGAGGA 2592

 288 uHisGlnAlaIleLeuSerLysArgLeuGluSerValGluSerAspIleA 305
2593 GCGCGCTGCGGAGCTGGCGAGCCAGCTGGAGGCCACTGCTGCTGCGAAGT 2642

 305 rgLeuGluGlnAsnThrMetAspMetThrArgIle.............. 316
2643 CGTCGGCGGAGCAGGACCGCGAGAACACGAGGGCCACGCTAGAGCAGCAG 2692

 317 ............AspAlaArgLysMetGlnMetThrGlyAspLeuIleMe 329
2693 CTTCGCGAATCCGAGGCGCGCGCTGCGGAGCTGGCGAGCCAGCTG..... 2737

 329 tLysAsnSerValThrValGlyGlyIleAlaGlyAlaSerGlyGlnTyrA 346
2738 .......................GAGGCCACTGCTGCTGCGAAGATGT 2762

 346 laAlaThrGlnGluArgSer.................GluGlnGlnIle 356
2763 CAGCGGAGCAGGACCGCGAGAACACGAGGGCCACGCTAGAGCAGCAGCTT 2812

 357 SerGlnValAsnAsnArgValAlaSerThrAlaSer.............. 368
2813 CGTGACTCCGAGGAGCGCGCTGCGGAGCTGGCGAGCCAGCTGGAGTCCAC 2862

 369 .....AspGluAlaArgGluSerSerArgLysS 378
2863 TACTGCTGCGAAGATGTCAGCGGAGCAGGACCGCGAGAGCACGAGG.... 2908

 378 erThrSerLeuIleGlnGluMetLeuLysThrMetGluSerIleAsnGln 394
2909 ..GCCACGCTAGAGCAGCAGCTTCGTGACTCCGAGGAGCGCGCTGCGGAG 2956

 395 SerLysAlaSerAlaLeuAlaAlaIleAlaGlyAsnIleArgAla*** 410
2957 CTGGCGAGCCAGCTGGAGTCCACTACTGCTGCGAAGATGTCAGCGGAG 3004
```

seq_name: N_Geneseq_36:T42166

seq_documentation_block:

```
ID   T42166 standard; DNA; 3319 BP.
AC   T42166;
DT   09-MAR-1997  (first entry)
DE   Leishmania chagasi K39 antigen.
KW   Leishmania chagasi; acidic ribosomal antigen; LcP0;
KW   epitope; K39; ss.
OS   Leishmania chagasi.
FH   Key             Location/Qualifiers
FT   cds             455. .3319
FT                   /*tag=  a
PN   WO9633414-A2.
PD   24-OCT-1996.
PF   19-APR-1996; U05472.
PR   21-APR-1995; US-428414.
PA   (CORI-) CORIXA CORP.
PI   Reed SG;
DR   WPI; 96-485884/48.
DR   P-PSDB; W03691.
PT   New Leishmania acidic ribosomal P-protein family poly:peptide - used
PT   to develop prods. for diagnosis, detection and protection against
PT   Leishmania infections
PS   Disclosure; Page 43-48; 76pp; English.
CC   Compounds including polypeptides that contain at least an epitope of
CC   the L. chagasi acidic ribosomal antigen LcP0 are useful in a variety
CC   of immunoassays for detecting Leishmania infection.  Portions of
CC   LcP0 (T42164) contg. at least the 17 C-terminal amino acids (T42165)
CC   have been found to generate a signal in an ELISA that is equivalent
CC   to that generated by the full length LcP0.   A combination
CC   polypeptide may also be used, comprising an LcP0 epitope along with
CC   an epitope derived from the Leishmania K39 antigen (T42166), pref.
CC   the K39 repeat unit antigen having the sequence given in W03690.
SQ   Sequence   3319 BP;    669 A;    969 C;    1196 G;    485 T;
```

```
25319 ATTACAAATGCAAATGATTTAAATCAAGCAATGACACAATTGAAACAAGG 25368

  107 .........ArgGluGlnGlnAlaGluValGlyLysPhePheAsp.... 118
25369 TATTGCAGATAAAGACCAAACTAAAGCAAATGGTAACTTTGTCAATGCTG 25418

  118 ..................................................... 118

25419 ATACTGATAAGCAAAATGCTTACAACAATGCGGTAGCACATGCTGAACAA 25468

  119 ...IleSerGlyMetSerSerSerAlaValAlaLeuLeuAlaAlaAlaAs 134
25469 ATAATTAGTGGTACACCAAATGCAAACGTGGATCCACAACAAGTGGCTCA 25518

  134 nThrLeuMetLeuThrLeuAsnGlnAlaAspSerLysLeuSerGlyLysL 151
25519 AGCGTTACAA...CAAGTGAATCAAGCTAAGGGTGATTTAAACGGTAACC 25565

  151 euSerLeu...ValSerPheAspAlaAlaLysThrThrAla......... 163
25566 ATAACTTACAAGTTGCTAAAGACAATGCAAATACAGCCATTGATCAGTTA 25615

  164 SerSerMetMetArgGluGlyMetAsnAlaLeuSerGlySerIleSerGl 180
25616 CCAAACTTAAATCAACCACAAAAAACAGCATTAAAAGACCAAGTGTCGCA 25665

  180 nSerAlaLeuGlnLeuGlyIleThrGlyValGlyAlaLysLeuGluTyrL 197
25666 TGCAGAACTTGTTACAGGTGTTAATGCTATT......AAGCAAAATGCTG 25709

  197 ysGlyLeuGlnAsnGluArgGlyAlaLeuLysHis.............. 208
25710 ATGCGTTAAATAATGCAATGGGTACATTGAAACAACAAATTCAAGCGAAC 25759

  209 .................................AsnAlaAlaLysIl 213
25760 AGTCAAGTACCACAGTCAGTTGACTTTACACAAGCGGATCAAGACAAACA 25809

  213 eAspLysLeuThrThrGluSerHisSerIleLysAsnValLeuAsnGlyG 230
25810 ACAAGCATATAACAATGCGGCTAACCAAGCGCAACAAATCGCAAATGGCA 25859

  230 lnAsnSerValLysLeuGlyAlaGluGlyVal...AspSerLeuLysSer 245
25860 TACCAACACCTGTATTGACGCCTGATACAGTAACACAAGCAGTGACAACT 25909

  246 LeuAsnMetLysLysThrGlyThrAspAlaThrLysAsnLeuAsnAspAl 262
25910 ATGAATCAAGCGAAAGATGCATTAAACGGTGATGAAAAATTAGCACAAGC 25959

  262 aThrLeuLysSerAsnAlaGlyThrSerAlaThrGluSerLeuGlyIleL 279
25960 GAAA.........................CAAGAAGCTTTAGCAAATC 25982

  279 ysAspSerAsnLysGlnIleSerProGluHisGlnAlaIleLeuSerLys 295
25983 TTGATACGTTACGCGATTTAAATCAACCACAACGTGATGCATTACGTAAC 26032

  296 ArgLeuGluSerValGluSerAspIleArgLeuGluGlnAsnThrMetAs 312
26033 CAAATCAATCAAGCACAAGCGTTAGCTACAGTTGAACAAACTAAACAAAA 26082

  312 pMetThrArgIleAspAlaArgLysMetGlnMetThrGlyAspLeuIleM 329
26083 TGCACAAAATGTGAATACAGCAATGAGTAACTTGAAACAAGGTATTGCAA 26132

  329 etLysAsnSerValThrValGly..............GlyIleAlaGly 340
26133 ACAAAGATACTGTCAAAGCAAGTGAGAACTATCATGATGCTGATGCCGAT 26182

  341 AlaSerGlyGlnTyrAlaAlaThrGlnGluArgSerGluGlnGlnIleSe 357

26183 AAGCAAACAGCATATACAAATGCAGTGTCTCAAGCGGAAGGTATTATCAA 26232

  357 rGlnValAsnAsnArgVal.....................AlaSerT 366
26233 TCAAACGACAAATCCAACGCTTAACCCAGATGAAATAACACGTGCATTAA 26282

  366 hrAlaSerAspGluAlaArgGluSerSerArgLysSerThrSerLeuIle 382
26283 CTCAAGTGACTGATGCTAAAAATGGCTTAAACGGTGAAGCTAAATTGGCA 26332

  383 GlnGluMetLeuLysThrMetGluSerIleAsn 393
26333 ACTGAAAAGCAAAATGCTAAAGATGCCGTAAGT 26365
```

seq_name: N_Geneseq_36:Q70152

seq_documentation_block:
ID Q70152 standard; cDNA to mRNA; 3319 BP.
AC Q70152;
DT 08-MAR-1995 (first entry)
DE K39 gene of Leishmania chagasi.
KW Parasite; leishmaniasis; infection; repeat unit; vaccine; antigen;
KW Leishmania chagasi; ss.
OS Leishmania chagasi.
FH Key Location/Qualifiers
FT cds 455. .3319
FT /*tag= a
FT /product= K39.
PN WO9416331-A.
PD 21-JUL-1994.
PF 10-JAN-1994; U00324.
PR 15-JAN-1993; US-006676.
PA (IASY-) IASYS CORP.
PI Reed SG;
DR WPI; 94-249402/30.
DR P-PSDB; R57365.
PT Diagnosis of Leishmaniasis - by determining the presence of
PT antibodies that bind to a K39 repeat unit antigen
PS Disclosure; Page 15-17; 28pp; English.
CC The K39 gene encodes a polypeptide which comprises a number of
CC repeated units (described in R57366). Detection of antibodies
CC directed against this repeated unit in a patients sample is
CC indicative of leishmaniasis. The antigenic repeat unit can
CC itself be used as a vaccine to protect against infection by a
CC leishmania parasite.
SQ Sequence 3319 BP; 669 A; 969 C; 1196 G; 485 T;

alignment_scores:
Quality: 109.50 Length: 516
Ratio: 0.429 Gaps: 22
Percent Similarity: 49.419 Percent Identity: 19.574

alignment_block:
US-09-068-804-6 x Q70152 ..

Align seg 1/1 to: Q70152 from: 1 to: 3319

```
    2 LeuIleSerAsnValGlyIleAsnProAlaAlaTyrLeuAsnAsnHisSe 18
1547 ATGATCGCGACTGTGAGCCCGAGCGCGCTGAACTAC.............. 1582

   18 rValGluAsnSerSerGlnThrAlaSerGlnSerValSerAlaLysAspI 35
1583 ....GAGGAGACGCTGAGCACGCTGCGGTACGCGTCGCGCGCGCGCGACA 1628

   35 leLeuAsnSerIleGlyIleSer............SerSerLysValSer 47
1629 TTGTGAATGTTGCGCAGGTGAACGAGGACCCGCGCGCACGGCGGATCCGC 1678

   48 AspLeu..............GlyLeuSerProThrLeuSerAlaProAl 59
1679 GAGCTGGAGGAGCAGATGGAGGACATGCGGCAGGCGATGGCTGGCGGCGA 1728
```

```
FT   misc_feature    8401. .8460
FT                   /*tag= e
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
FT                          are included to maintain the nucleotide numbering
FT                          given in the specification for this DNA sequence"
FT   misc_feature    10201. .10260
FT                   /*tag= f
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
FT                          are included to maintain the nucleotide numbering
FT                          given in the specification for this DNA sequence"
FT   misc_feature    12001. .12060
FT                   /*tag= g
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
FT                          are included to maintain the nucleotide numbering
FT                          given in the specification for this DNA sequence"
FT   misc_feature    13801. .13860
FT                   /*tag= h
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
FT                          are included to maintain the nucleotide numbering
FT                          given in the specification for this DNA sequence"
FT   misc_feature    15601. .15660
FT                   /*tag= i
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
FT                          are included to maintain the nucleotide numbering
FT                          given in the specification for this DNA sequence"
FT   misc_feature    17401. .17460
FT                   /*tag= j
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
FT                          are included to maintain the nucleotide numbering
FT                          given in the specification for this DNA sequence"
FT   misc_feature    19201. .19260
FT                   /*tag= k
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
FT                          are included to maintain the nucleotide numbering
FT                          given in the specification for this DNA sequence"
FT   misc_feature    21001. .21060
FT                   /*tag= l
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
FT                          are included to maintain the nucleotide numbering
FT                          given in the specification for this DNA sequence"
FT   misc_feature    22801. .22860
FT                   /*tag= m
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
FT                          are included to maintain the nucleotide numbering
FT                          given in the specification for this DNA sequence"
FT   misc_feature    24601. .24660
FT                   /*tag= n
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
FT                          are included to maintain the nucleotide numbering
FT                          given in the specification for this DNA sequence"
FT   misc_feature    26401. .26460
FT                   /*tag= o
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
FT                          are included to maintain the nucleotide numbering
FT                          given in the specification for this DNA sequence"
FT   misc_feature    28201. .28260
FT                   /*tag= p
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
FT                          are included to maintain the nucleotide numbering
FT                          given in the specification for this DNA sequence"
FT   misc_feature    30001. .30060
FT                   /*tag= q
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
FT                          are included to maintain the nucleotide numbering
FT                          given in the specification for this DNA sequence"
PN   EP-786519-A2.
PD   30-JUL-1997.
PF   07-JAN-1997; 100117.
PR   05-JAN-1996; US-009861.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI   Rosen CA;
DR   WPI; 97-374922/35.
PT   Polynucleotide(s) and proteins derived from Staphylococcus aureus -
PT   stored on computer readable medium and used in the production of
PT   anti-S.aureus vaccines
PS   Claim 1; Page 452-469; 3271pp; English.
CC   This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC   of the invention. The DNA sequences are recorded on a computer readable
CC   medium, preferably selected from a floppy or hard disk, random access
CC   memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC   the S.aureus DNA sequences allows putative functions to be assigned so
CC   that protein-encoding or regulatory regions of commercial, therapeutic or
CC   industrial importance can be obtained. Specifically, sequences which are
CC   likely to encode antigens have been identified and these polypeptides can
CC   be used in a vaccine composition against S.aureus infection. The
CC   polypeptides can also be used in a kit for the immunodetection of
CC   S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC   including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC   skin and surgical wound infections, scalded skin syndrome, toxic shock
CC   syndrome, etc. Organisms transformed with the DNA sequences can be used
CC   for recombinant production of the polypeptides. The new DNA sequences
CC   (and their fragments) are useful as primers or probes for isolating
CC   homologues of any of the S.aureus DNA sequences contained on the
CC   computer readable medium.
SQ   Sequence   31096 BP;    11857 A;    5243 C;    5477 G;    7488 T;


alignment_scores:
           Quality:  127.50              Length:    461
             Ratio:  0.500                Gaps:     15
 Percent Similarity:  55.315   Percent Identity: 17.570

alignment_block:
 US-09-068-804-6 x V74370   ..

 Align seg 1/1  to: V74370  from: 1  to: 31096

      15 AsnAsnHisSerValGluAsnSerSerGlnThrAlaSerGlnSerValSe 31
         |||:::::::::::::::::|||:::      :::   ::::::::::::
   25040 AATGCTCAAAATCTGAATAATGCTATGACTAACTTAAACAATGCATTACA 25089

      31 rAlaLys...AspIleLeuAsnSerIleGlyIleSerSerSerLysValS 47
          :::|||   ::::::|||||||||||||:::::::::::::::      :
   25090 AGATAAAACTGAGACATTAAATAGCATTAACTTTACTGATGCAGATCAAG 25139

      47 erAspLeuGlyLeuSerProThrLeuSerAlaProAlaProGlyValLeu 63
         ::      :::      ::::::      :::   |||   |||:::|||
   25140 CTAAGAAAGATGCTTATACTAATGCGGTTTCACATGCAGAAGGTATTTTA 25189

      64 ThrGlnThrProGlyThrIleThrSerPheLeuLysAlaSerIleGlnAs 80
         :::::::::   |||:::   :::|||                      ::
   25190 TCTAAAGCAAATGGCAGCAATGCAAGT.....................CA 25218

      80 nThrAspMetAsnGlnAspLeu..............................  87
         :|||::::::::::|||::::::
   25219 AACTGAAGTGGAACAAGCGATGCAACGTGTGAACGAAGCGAAACAAGCAT 25268

      88 ............................AsnAlaLeuAlaAsnAsnVal  94
                                     ::::::   |||:::::|||
   25269 TGAATGGTAATGACAATGTACAACGTGCAAAAGATGCAGCGAAACAAGTG 25318

      95 ThrThrLysAlaAsnGluValValGln......ThrGlnLeu........ 106
```

```
          Quality: 1607.00              Length:    410
            Ratio:  3.939                Gaps:      1
Percent Similarity:  99.512   Percent Identity: 99.268

alignment_block:
US-09-068-804-6 x T67035   ..

Align seg 1/1  to: T67035  from: 1  to: 3441

       1 MetLeuIleSerAsnValGlyIleAsnProAlaAlaTyrLeuAsnAsnHi 17
         ||||||||||:::|||||||||||||||||||||||||||||||||||||
     898 ATGTTAATTACTAATGTGGGAATAAATCCCGCCGCTTATTTAAATAATCA 947

      17 sSerValGluAsnSerSerGlnThrAlaSerGlnSerValSerAlaLysA 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     948 TTCTGTTGAGAATAGTTCACAGACAGCTTCGCAATCCGTTAGCGCTAAAG 997

      34 spIleLeuAsnSerIleGlyIleSerSerSerLysValSerAspLeuGly 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     998 ATATTCTGAATAGTATTGGTATTAGCAGCAGTAAAGTCAGTGACCTGGGG 1047

      51 LeuSerProThrLeuSerAlaProAlaProGlyValLeuThrGlnThrPr 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1048 TTGAGTCCTACACTGAGCGCGCCTGCGCCAGGGGTATTAACGCAAACCCC 1097

      67 oGlyThrIleThrSerPheLeuLysAlaSerIleGlnAsnThrAspMetA 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1098 CGGAACGATCACGTCCTTTTTAAAAGCCAGTATTCAAAATACCGACATGA 1147

      84 snGlnAspLeuAsnAlaLeuAlaAsnAsnValThrThrLysAlaAsnGlu 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1148 ATCAGGATTTGAATGCTCTGGCAAATAATGTCACGACTAAAGCGAATGAG 1197

     101 ValValGlnThrGlnLeuArgGluGlnGlnAlaGluValGlyLysPhePh 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1198 GTTGTGCAAACCCAGTTACGCGAGCAGCAGGCAGAAGTCGGAAAGTTTTT 1247

     117 eAspIleSerGlyMetSerSerSerAlaValAlaLeuLeuAlaAlaAlaA 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1248 TGATATTAGCGGAATGTCTTCCAGTGCCGTTGCGCTGTTGGCTGCCGCGA 1297

     134 snThrLeuMetLeuThrLeuAsnGlnAlaAspSerLysLeuSerGlyLys 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1298 ATACGTTAATGCTGACGTTGAACCAGGCTGATAGCAAACTGTCTGGTAAG 1347

     151 LeuSerLeuValSerPheAspAlaAlaLysThrThrAlaSerSerMetMe 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1348 TTGTCATTAGTCAGTTTTGATGCAGCTAAAACGACGGCAAGCTCCATGAT 1397

     167 tArgGluGlyMetAsnAlaLeuSerGlySerIleSerGlnSerAlaLeuG 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1398 GCGCGAAGGGATGAATGCGTTGTCCGGTAGTATTTCCCAGAGCGCGCTTC 1447

     184 lnLeuGlyIleThrGlyValGlyAlaLysLeuGluTyrLysGlyLeuGln 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1448 AGTTGGGGATCACTGGCGTGGGCGCCAAACTGGAATATAAGGGGCTGCAG 1497

     201 AsnGluArgGlyAlaLeuLysHisAsnAla.AlaLysIleAspLysLeuT 217
         |||||||||||||||||||||||||||||| ||||||||||||||||||||
    1498 AATGAAAGAGGCGCGCTTAAACATAATGCCCGCGAAGATCGATAAACTGA 1547

     217 hrThrGluSerHisSerIleLysAsnValLeuAsnGlyGlnAsnSerVal 233
         |||||||||||||||||||||||||||||   ||||||||||||||||||
    1548 CCACTGAAAGCCACAGTATTAAAAACGTGCG.AACGGGCAGAATAGCGTC 1596

     234 LysLeuGlyAlaGluGlyValAspSerLeuLysSerLeuAsnMetLysLy 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1597 AAACTCGGTGCTGAAGGCGTCGATTCTCTGAAATCGTTAAATATGAAGAA 1646

     250 sThrGlyThrAspAlaThrLysAsnLeuAsnAspAlaThrLeuLysSerA 267
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1647 AACCGGTACCGATGCGACGAAAAATCTTAATGATGCGACGCTTAAATCTA 1696

     267 snAlaGlyThrSerAlaThrGluSerLeuGlyIleLysAspSerAsnLys 283
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1697 ATGCCGGAACCAGCGCCACGGAAAGTCTGGGTATTAAAGACAGTAATAAA 1746

     284 GlnIleSerProGluHisGlnAlaIleLeuSerLysArgLeuGluSerVa 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1747 CAAATCTCCCCTGAACATCAGGCTATTCTGTCGAAACGTCTTGAGTCTGT 1796

     300 lGluSerAspIleArgLeuGluGlnAsnThrMetAspMetThrArgIleA 317
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1797 CGAATCCGATATTCGTCTTGAGCAGAATACCATGGATATGACCCGAATCG 1846

     317 spAlaArgLysMetGlnMetThrGlyAspLeuIleMetLysAsnSerVal 333
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1847 ATGCGCGCAAGATGCAGATGACGGGCGATCTGATTATGAAGAACTCGGTC 1896

     334 ThrValGlyGlyIleAlaGlyAlaSerGlyGlnTyrAlaAlaThrGlnGl 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1897 ACGGTCGGTGGTATTGCAGGGGCGTCCGGGCAGTACGCCGCTACTCAGGA 1946

     350 uArgSerGluGlnGlnIleSerGlnValAsnAsnArgValAlaSerThrA 367
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1947 ACGTTCCGAGCAGCAAATTAGCCAGGTGAATAACCGGGTTGCCAGCACCG 1996

     367 laSerAspGluAlaArgGluSerSerArgLysSerThrSerLeuIleGln 383
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1997 CATCGGACGAAGCCCGTGAAAGTTCACGTAAATCGACCAGCCTGATTCAG 2046

     384 GluMetLeuLysThrMetGluSerIleAsnGlnSerLysAlaSerAlaLe 400
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2047 GAAATGCTGAAAACAATGGAGAGCATTAACCAGTCGAAAGCATCCGCACT 2096

     400 uAlaAlaIleAlaGlyAsnIleArgAla 409
         |||||||||||||||||||||||||
    2097 CGCTGCTATCGCAGGCAATATTCGCGCT 2124

seq_name: N_Geneseq_36:V74370

seq_documentation_block:
ID   V74370 standard; DNA; 31096 BP.
AC   V74370;
DT   16-MAR-1999  (first entry)
DE   Staphylococcus aureus contig SEQ ID #59.
KW   Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW   cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW   skin infection; surgical wound infection; scalded skin syndrome;
KW   toxic shock syndrome; ds.
OS   Staphylococcus aureus.
FH   Key             Location/Qualifiers
FT   misc_feature    1201. .1260
FT                   /*tag= a
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
FT                          are included to maintain the nucleotide numbering
FT                          given in the specification for this DNA sequence"
FT   misc_feature    3001. .3060
FT                   /*tag= b
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
FT                          are included to maintain the nucleotide numbering
FT                          given in the specification for this DNA sequence"
FT   misc_feature    4801. .4860
FT                   /*tag= c
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
FT                          are included to maintain the nucleotide numbering
FT                          given in the specification for this DNA sequence"
FT   misc_feature    6601. .6660
FT                   /*tag= d
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
FT                          are included to maintain the nucleotide numbering
FT                          given in the specification for this DNA sequence"
```

```
 101 ATATTCTGAATAGTATTGGTATTAGCAGCAGTAAAGTCAGTGACCTGGGG 150

  51 LeuSerProThrLeuSerAlaProAlaProGlyValLeuThrGlnThrPr 67
     |||||||||||||||||||||||||||||||||||||||||||||||||||
 151 TTGAGTCCTACACTGAGCGCGCCTGCGCCAGGGGTATTAACGCAAACCCC 200

  67 oGlyThrIleThrSerPheLeuLysAlaSerIleGlnAsnThrAspMetA 84
     |||||||||||||||||||||||||||||||||||||||||||||||||||
 201 CGGAACGATCACGTCCTTTTTAAAAGCCAGTATTCAAAATACCGACATGA 250

  84 snGlnAspLeuAsnAlaLeuAlaAsnAsnValThrThrLysAlaAsnGlu 100
     |||||||||||||||||||||||||||||||||||||||||||||||||||
 251 ATCAGGATTTGAATGCTCTGGCAAATAATGTCACGACTAAAGCGAATGAG 300

 101 ValValGlnThrGlnLeuArgGluGlnGlnAlaGluValGlyLysPhePh 117
     |||||||||||||||||||||||||||||||||||||||||||||||||||
 301 GTTGTGCAAACCCAGTTACGCGAGCAGCAGGCAGAAGTCGGAAAGTTTTT 350

 117 eAspIleSerGlyMetSerSerSerAlaValAlaLeuLeuAlaAlaAlaA 134
     |||||||||||||||||||||||||||||||||||||||||||||||||||
 351 TGATATTAGCGGAATGTCTTCCAGTGCCGTTGCGCTGTTGGCTGCCGCGA 400

 134 snThrLeuMetLeuThrLeuAsnGlnAlaAspSerLysLeuSerGlyLys 150
     |||||||||||||||||||||||||||||||||||||||||||||||||||
 401 ATACGTTAATGCTGACGTTGAACCAGGCTGATAGCAAACTGTCTGGTAAG 450

 151 LeuSerLeuValSerPheAspAlaAlaLysThrThrAlaSerSerMetMe 167
     |||||||||||||||||||||||||||||||||||||||||||||||||||
 451 TTGTCATTAGTCAGTTTTGATGCAGCTAAAACGACGGCAAGCTCCATGAT 500

 167 tArgGluGlyMetAsnAlaLeuSerGlySerIleSerGlnSerAlaLeuG 184
     |||||||||||||||||||||||||||||||||||||||||||||||||||
 501 GCGCGAAGGGATGAATGCGTTGTCCGGTAGTATTTCCCAGAGCGCGCTTC 550

 184 lnLeuGlyIleThrGlyValGlyAlaLysLeuGluTyrLysGlyLeuGln 200
     |||||||||||||||||||||||||||||||||||||||||||||||||||
 551 AGTTGGGGATCACTGGCGTGGGCGCCAAACTGGAATATAAGGGGCTGCAG 600

 201 AsnGluArgGlyAlaLeuLysHisAsnAlaAlaLysIleAspLysLeuTh 217
     |||||||||||||||||||||||||||||||||||||||||||||||||||
 601 AATGAAAGAGGCGCGCTTAAACATAATGCCGCGAAGATCGATAAACTGAC 650

 217 rThrGluSerHisSerIleLysAsnValLeuAsnGlyGlnAsnSerValL 234
     |||||||||||||||||||||||||||||||||||||||||||||||||||
 651 CACTGAAAGCCACAGTATTAAAAACGTGCTGAACGGGCAGAATAGCGTCA 700

 234 ysLeuGlyAlaGluGlyValAspSerLeuLysSerLeuAsnMetLysLys 250
     |||||||||||||||||||||||||||||||||||||||||||||||||||
 701 AACTCGGTGCTGAAGGCGTCGATTCTCTGAAATCGTTAAATATGAAGAAA 750

 251 ThrGlyThrAspAlaThrLysAsnLeuAsnAspAlaThrLeuLysSerAs 267
     |||||||||||||||||||||||||||||||||||||||||||||||||||
 751 ACCGGTACCGATGCGACGAAAAATCTTAATGATGCGACGCTTAAATCTAA 800

 267 nAlaGlyThrSerAlaThrGluSerLeuGlyIleLysAspSerAsnLysG 284
     |||||||||||||||||||||||||||||||||||||||||||||||||||
 801 TGCCGGAACCAGCGCCACGGAAAGTCTGGGTATTAAAGACAGTAATAAAC 850

 284 lnIleSerProGluHisGlnAlaIleLeuSerLysArgLeuGluSerVal 300
     |||||||||||||||||||||||||||||||||||||||||||||||||||
 851 AAATCTCCCCTGAACATCAGGCTATTCTGTCGAAACGTCTTGAGTCTGTC 900

 301 GluSerAspIleArgLeuGluGlnAsnThrMetAspMetThrArgIleAs 317
     |||||||||||||||||||||||||||||||||||||||||||||||||||
 901 GAATCCGATATTCGTCTTGAGCAGAATACCATGGATATGACCCGAATCGA 950

 317 pAlaArgLysMetGlnMetThrGlyAspLeuIleMetLysAsnSerValT 334
     |||||||||||||||||||||||||||||||||||||||||||||||||||
 951 TGCGCGCAAGATGCAGATGACGGGCGATCTGATTATGAAGAACTCGGTCA 1000

 334 hrValGlyGlyIleAlaGlyAlaSerGlyGlnTyrAlaAlaThrGlnGlu 350
     |||||||||||||||||||||||||||||||||||||||||||||||||||
1001 CGGTCGGTGGTATTGCAGGGGCGTCCGGGCAGTACGCCGCTACTCAGGAA 1050
```

```
 351 ArgSerGluGlnGlnIleSerGlnValAsnAsnArgValAlaSerThrAl 367
     |||||||||||||||||||||||||||||||||||||||||||||||||||
1051 CGTTCCGAGCAGCAAATTAGCCAGGTGAATAACCGGGTTGCCAGCACCGC 1100

 367 aSerAspGluAlaArgGluSerSerArgLysSerThrSerLeuIleGlnG 384
     |||||||||||||||||||||||||||||||||||||||||||||||||||
1101 ATCGGACGAAGCCCGTGAAAGTTCACGTAAATCGACCAGCCTGATTCAGG 1150

 384 luMetLeuLysThrMetGluSerIleAsnGlnSerLysAlaSerAlaLeu 400
     |||||||||||||||||||||||||||||||||||||||||||||||||||
1151 AAATGCTGAAAACAATGGAGAGCATTAACCAGTCGAAAGCATCCGCACTC 1200

 401 AlaAlaIleAlaGlyAsnIleArgAla 409
     |||||||||||||||||||||||||
1201 GCTGCTATCGCAGGCAATATTCGCGCT 1227
```

seq_name: N_Geneseq_36:T67035

seq_documentation_block:

ID T67035 standard; DNA; 3441 BP.
AC T67035;
DT 04-AUG-1997 (first entry)
DE Salmonella virulence factor SspBCDA DNA region.
KW Salmonella secreted protein; SspB; SspC; SspD; SspA;
KW bacterial-mediated endocytosis; diagnosis; therapy; vaccine;
KW attenuation; virulence; ss.
OS Salmonella typhimurium.

FH	Key	Location/Qualifiers
FT	cds	1. .870
FT		/*tag= a
FT		/product= SspB (incomplete)
FT	cds	898. .2127
FT		/*tag= b
FT		/product= SspC
FT	cds	2198. .3228
FT		/*tag= c
FT		/product= SspD
FT	cds	3247. .3441
FT		/*tag= d
FT		/product= SspA (incomplete)

PN WO9718225-A1.
PD 22-MAY-1997.
PF 14-NOV-1996; U18504.
PR 14-NOV-1995; US-006733.
PA (GEHO) GEN HOSPITAL CORP.
PI Miller SI;
DR WPI; 97-289217/26.
DR P-PSDB; W15268-71.
PT New isolated Salmonella secreted proteins and related genes - used
PT to develop products for the detection, treatment or prevention of
PT Salmonella infections
PS Claim 11; Fig 28; 95pp; English.
CC A DNA molecule (T67035) comprises the SspBCDA locus (truncated at
CC the 5' and 3' ends) including the partial sspB, sspC, sspD and
CC partial sspA sequences (see also T67028-31) coding for Salmonella
CC typhimurium secreted proteins SspB, SspC, SspD and SspA (W15268-
CC 70). To identify the ssp genes, S. typhimurium mutants were
CC investigated in invasion assays on HEp-2 epithelial cells. 2
CC Ssp mutants with transposon insertions located between spaT and
CC prgH (see also T67036) were identified. One mutant lacked Ssp of
CC 87, 42 and 36 kDa. The invasiveness of this mutant implicated the
CC 42 and 36 kDa Ssp in invasion. DNA sequencing of this region
CC identified the 4 ssp (partial) genes. The ssp genes can be used in
CC methods for diagnosing Salmonella infection and in methods of
CC targetting an antigen to epithelial cells and of inducing a
CC cytotoxic T cell immune response in a mammal. Ssp gene mutation
CC can be used to create attenuated Salmonella strains useful as live
CC vaccines.
SQ Sequence 3441 BP; 978 A; 761 C; 925 G; 777 T;

alignment_scores:

```
OM of: US-09-068-804-6 to: N_Geneseq_36:*   out_format : pfs

Date: Oct 5, 1999  8:02 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn2_1/USPTO_spool/US09068804/runat_05101999_145135_10341/app_query.fasta.1
 -DB=N_Geneseq_36 -QFMT=fastap -SUFFIX=rng -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=pam150
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=escore
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09068804
 -NCPU=6 -ICPU=3 -WAIT -THREADS=1

Search information block:
Query: US-09-068-804-6
Query length: 410
Database: N_Geneseq_36:*
Database sequences: 311585
Database length: 125096042
Search time (sec): 156.120000
```

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ..
N_Geneseq_36:T67029	+	1649.00	3522.53	1.3e-188	1230	! Salmonella secreted protein Ss
N_Geneseq_36:T67035	+	1607.00	3421.36	5.7e-183	3441	! Salmonella virulence factor Ss
N_Geneseq_36:V74370	+	127.50	216.06	0.0002	31096	! Staphylococcus aureus contig
N_Geneseq_36:Q70152	+	109.50	200.92	0.0014	3319	! K39 gene of Leishmania chagasi
N_Geneseq_36:T42166	+	109.50	200.92	0.0014	3319	! Leishmania chagasi K39 antigen
N_Geneseq_36:V31200	+	118.00	198.96	0.0018	22671	! E. coli J96 pathogenicity isl
N_Geneseq_36:V31271	-	109.00	188.97	0.0063	9319	! E. coli J96 pathogenicity isla
N_Geneseq_36:V41346	+	102.00	187.07	0.0081	2673	! M. catarrhalis strain TTA24 Us
N_Geneseq_36:T58751	+	104.50	186.33	0.0089	4779	! Nucleolar/endosomal auto-antig
N_Geneseq_36:T14618	+	99.00	183.96	0.0120	1947	! DHFR/C. pneumoniae antigen fus
N_Geneseq_36:Q49506	+	102.50	181.31	0.0169	5116	! Sequence encoding high molecul
N_Geneseq_36:Q72293	+	102.50	181.31	0.0169	5116	! Haemophilus high molecular wei
N_Geneseq_36:V22834	+	102.50	175.44	0.0359	8930	! Haemophilus paragallinarum ant
N_Geneseq_36:T90996	+	102.50	175.16	0.0372	9171	! Non-typeable Haemophilus influ
N_Geneseq_36:Q49508	+	102.50	175.10	0.0375	9220	! Gene cluster for high molecula
N_Geneseq_36:T90994	+	99.50	174.86	0.0387	5116	! Non-typeable Haemophilus high
N_Geneseq_36:V34610	-	91.00	174.57	0.0401	927	! M. vaccae antigen GV-27B encodi
N_Geneseq_36:X20338	-	91.50	174.54	0.0403	1030	! Borrelia burgdorferi polynucle
N_Geneseq_36:T14612	+	93.00	174.06	0.0428	1464	! Chlamydia pneumoniae polypepti
N_Geneseq_36:X32494	+	92.00	170.96	0.0637	1602	! Seq ID No:9 used in a method f
N_Geneseq_36:X00225	+	91.50	170.16	0.0706	1560	! Clostridium chauvoei blackleg
N_Geneseq_36:Q99456	+	99.00	170.00	0.0721	7326	! Sequence encoding CREB binding
N_Geneseq_36:V10093	+	99.00	170.00	0.0721	7326	! Cellular transcriptional facto
N_Geneseq_36:V34608	-	91.00	169.03	0.0817	1569	! M. vaccae antigen GV-27 encodi
N_Geneseq_36:T68154	+	88.00	167.52	0.0990	981	! H. pylori secreted or periplasm
N_Geneseq_36:Q49510	+	95.00	167.04	0.1053	4286	! Partial sequence encoding high
N_Geneseq_36:Q72295	+	95.00	167.04	0.1054	4287	! Haemophilus high molecular wei
N_Geneseq_36:Q49511	+	95.00	166.07	0.1194	4702	! Partial sequence encoding high
N_Geneseq_36:Q72296	+	95.00	166.07	0.1194	4702	! Haemophilus high molecular wei
N_Geneseq_36:T90992	+	95.00	165.86	0.1226	4794	! Non-typeable Haemophilus high
N_Geneseq_36:Q45214	+	86.00	164.60	0.1442	861	! Recombinant M24-M5-M6 gene. New
N_Geneseq_36:T84162	-	92.50	164.15	0.1527	3386	! DNA encoding a lysostaphin pre
N_Geneseq_36:V53504	-	92.50	164.15	0.1527	3386	! DNA encoding a Staphylococcus
N_Geneseq_36:T74034	+	94.00	164.10	0.1537	4621	! Mouse male-enhanced antigen-2
N_Geneseq_36:X04132	+	94.00	164.10	0.1537	4621	! Mouse male enhanced antigen 2
N_Geneseq_36:V69792	+	92.00	163.52	0.1655	3246	! Candida albicans CaNik1 gene i
N_Geneseq_36:X40015	+	90.00	163.24	0.1715	2215	! Prostate cancer associated gen
N_Geneseq_36:V21060	+	92.50	162.72	0.1835	3879	! cDNA encoding the a tumour sup
N_Geneseq_36:V29009	+	92.50	162.72	0.1835	3879	! Human polyhomeotic 1 (hph1) on
N_Geneseq_36:T90993	+	93.50	162.62	0.1858	4803	! Non-typeable Haemophilus high
N_Geneseq_36:V41347	+	91.00	161.22	0.2223	3292	! M. catarrhalis strain TTA37 Us
N_Geneseq_36:T37177	+	87.50	161.07	0.2267	1635	! LKT-GnRH fusion of pCB111. Gon
N_Geneseq_36:V61532	+	87.50	161.07	0.2267	1635	! Nucleic acid encoding LKT-G
N_Geneseq_36:V30458_0	-	108.00	160.81	0.2343	110000	! Rhizobium species plasmi
N_Geneseq_36:V30458_1	-	108.00	160.81	0.2343	110000	! Continuation (2 of 6) of

```
seq_name: N_Geneseq_36:T67029

seq_documentation_block:
ID   T67029 standard; DNA; 1230 BP.
AC   T67029;
DT   04-AUG-1997  (first entry)
DE   Salmonella secreted protein SspC DNA.
KW   Salmonella secreted protein; SspC; bacterial-mediated endocytosis;
KW   diagnosis; therapy; vaccine; attenuation; virulence; ss.
OS   Salmonella typhimurium.
FH   Key             Location/Qualifiers
FT   cds             1. .1230
FT                   /*tag= a
FT                   /product= SspC
FT                   /note= "the SspC amino acid sequence deduced
FT                   from the coding sequence does not fully
FT                   correspond to the SspC sequence given in
FT                   Fig 13"
PN   WO9718225-A1.
PD   22-MAY-1997.
PF   14-NOV-1996; U18504.
PR   14-NOV-1995; US-006733.
PA   (GEHO ) GEN HOSPITAL CORP.
PI   Miller SI;
DR   WPI; 97-289217/26.
DR   P-PSDB; W15269.
PT   New isolated Salmonella secreted proteins and related genes - used
PT   to develop products for the detection, treatment or prevention of
PT   Salmonella infections
PS   Claim 5; Fig 20; 95pp; English.
CC   A DNA molecule (T67029) codes for Salmonella secreted protein SspC
CC   (W15269), a virulence factor involved in bacterial mediated
CC   endocytosis, which is able to translocate into human epithelial
CC   cells.  To identify Ssp genes, S. typhimurium mutants were
CC   investigated in invasion assays on HEp-2 epithelial cells.  2
CC   Ssp mutants with transposon insertions located between spaT and
CC   prgH were identified.  Of these, one mutant lacked Ssp of 87,
CC   42 and 36 kDa.  The invasiveness of this mutant implicated the
CC   42 and 36 kDa Ssp in invasion.  DNA sequencing of this region (see
CC   also T67035) identified 2 complete and 2 partial open reading
CC   frames, designated sspB, sspC (T67029), sspD (T67030) and sspA
CC   (T67031).  These nucleic acids can be used in methods for diagnosing
CC   Salmonella infection and in methods of targetting an antigen to
CC   epithelial cells and of inducing a cytotoxic T cell immune response
CC   in a mammal.  Ssp gene mutation can be used to create attenuated
CC   Salmonella strains useful as live vaccines.
SQ   Sequence   1230 BP;    362 A;    270 C;    317 G;    281 T;


alignment_scores:
           Quality: 1649.00              Length:    409
             Ratio:  4.032                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-068-804-6 x T67029   ..

 Align seg 1/1  to: T67029  from: 1  to: 1230

       1 MetLeuIleSerAsnValGlyIleAsnProAlaAlaTyrLeuAsnAsnHi 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 ATGTTAATTAGTAATGTGGGAATAAATCCCGCCGCTTATTTAAATAATCA 50

      17 sSerValGluAsnSerSerGlnThrAlaSerGlnSerValSerAlaLysA 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 TTCTGTTGAGAATAGTTCACAGACAGCTTCGCAATCCGTTAGCGCTAAAG 100

      34 spIleLeuAsnSerIleGlyIleSerSerSerLysValSerAspLeuGly 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
  15 AsnAsnHisSerValGluAsnSerSerGlnThrAlaSerGlnSerValSe 31
     ||||||:::::::   ::::::|||:::::::::||||:::   ::::::::
3616 AATAACAACACTGAAGATAGCAGTGACAATAATGCCGGCTTAACTATCGA 3665

  31 rAlaLysAspIleLeuAsnSerIleGlyIleSerSerSerLysValSerA 48
     :||||||:::::::      :::   :::|||:::|||   |||:::   :
3666 TGCAAAAAATGTAACAGTAAACAACAATATTACTTCTCACAAAGCAGTGA 3715

  48 spLeuGly..................................... 50
     ::::::::
3716 GCATCTCTGCGACAAGTGGAGAAATTACCACTAAAACAGGTACAACCATT 3765

  51 ..............................................Le 51
                                                    ||
3766 AACGCAACCACTGGTAACGTGGAGATAACCGCTCAAACAGGTAGTATCCT 3815

  51 uSerProThrLeuSerAlaProAlaProGlyValLeuThrGlnThrProG 68
     |:::   :::   |||:::::::::::::  :::|||||||   |||   |
3816 AGGTGGAATTGAGTCCAGCTCTGGCTCTGTAACACTTACTGCAACCGAGG 3865

  68 ly......ThrIleThrSerPheLeuLysAlaSerIleGlnAsnThrAsp 82
     ||      ::::::::::::::      :::::::::      |||:::
3866 GCGCTCTTGCTGTAAGCAATATTTCGGGCAACACCGTTACTGTTACTGCA 3915

  83 MetAsnGlnAspLeuAsnAlaLeuAlaAsnAsnValThrThrLysAlaAs 99
        :::   :::|||::::::||||||:::::::   |||:::|||:::::
3916 AATAGCGGTGCATTAACCACTTTGGCAGGCTCT...ACAATTAAAGGAAC 3962

  99 nGluValValGlnThrGlnLeuArgGluGlnGlnAlaGluValGlyLysP 116
     :|||   |||   |||           |||   :::::::::|||
3963 CGAGAGTGTAACCACT......TCAAGTCAATCAGGCGATATCGGC.... 4002

 116 hePheAspIleSerGlyMetSerSerSerAlaValAlaLeuLeuAlaAla 132
               :::   ::::||:::::::::|||::::::   |||:::
4003 ..........GGTACGATTTCTGGTGGCACAGTAGAGGTTAAAGCAACC 4041

 133 AlaAsnThrLeuMetLeuThrLeuAsnGlnAlaAspSerLysLeuSerGl 149
     ::::::         ||||||   :::|||::::::||||||:::   ::
4042 GAAAGT.........TTAACC...ACTCAATCCAATTCAAAAATTAAAGC 4079

 149 yLysLeuSerLeuValSerPheAspAlaAlaLysThrThrAlaSerSerM 166
     ::::   :::            :::|||:::  |||:::|||::::::
4080 AACAACAGGC............GAGGCTAACGTAACAAGTGCAACAGGTA 4117

 166 etMetArgGluGlyMetAsnAlaLeuSerGlySerIleSerGlnSerAla 182
       :::         :::::::::   :::   :::::::::   :::|||
4118 CAATTGGTGGTACGATTTCCGGTAATACGGTAAATGTTACGGCAAACGCT 4167

 183 LeuGlnLeuGlyIleThrGlyValGlyAlaLysLeuGluTyrLysGlyLe 199
        :::|||   :::   |||   ||||||   ::::::
4168 GGCGATTTAACAGTT...GGGAATGGCGCAGAAATTAAT.........GC 4205

 199 uGlnAsnGluArgGlyAlaLeuLysHisAsnAlaAlaLysIleAspLysL 216
        :::      ::::::|||:::   :::::::::        ||||
4206 GACAGAAGGAGCTGCAACCTTAACTACATCATCGGGC.........AAAT 4246

 216 euThrThrGluSerHisSerIleLysAsnValLeuAsnGlyGlnAsnSer 232
     ||||||||||:::   |||      ::::::   ::::::::::::
4247 TAACTACCGAAGCTAGTTCA......CACATTACTTCAGCCAAGGGTCAG 4290

 233 ValLysLeuGlyAlaGluGlyValAspSerLeuLysSerLeuAsnMetLy 249
     |||:::|||:::|||::::::   :::          |||:::|||
4291 GTAAATCTTTCAGCTCAGGATGGTAGCGTTGCAGGAAGTATTAATGCCGC 4340

 249 sLysThrGlyThrAspAlaThrLysAsnLeuAsnAspAlaThrLeuLysS 266
      ::::::      ::::::|||   :::|||      :::|||:::|||:
4341 CAATGTGACACTAAATACTACAGGCACTTTA......ACTACCGTGAAGG 4384

 266 erAsnAlaGlyThrSerAlaThrGluSerLeuGlyIleLysAspSerAsn 282
     ::::::::   ::::::::::::   :::|||   |||::::::   :::
4385 GTTCAAACATTAATGCAACCAGCGGTACCTTGGTTATTAACGCAAAAGAC 4434

 283 LysGlnIleSerProGluHisGlnAlaIleLeuSerLysArgLeuGluSe 299
        ::::::         :::   |||   |||::::::::::
4435 GCTGAGCTA.........AATGGCGCAGCATTGGGTAACCACACAGTGGT 4475

 299 rValGluSerAspIleArgLeuGluGlnAsnThrMetAspMetThrArgI 316
        :::::::::                ::::::::::::   |||
4476 AAATGCAACCAACGCAAATGGCTCCGGCAGCGTAATCGCGACAACC.... 4521

 316 leAspAlaArgLysMetGlnMetThrGlyAspLeuIleMetLysAsnSer 332
        :::   ::::::::::::|||||||||||||||||      |||:::
4522 .....TCAAGCAGAGTGAACATCACTGGGGATTTAATCACAATAAATGGA 4566

 333 ValThrVal 335
     :::::::::
4567 TTAAATATC 4575
```

```
  31 rAlaLysAspIleLeuAsnSerIleGlyIleSerSerSerLysValSerA 48
     :||||||::::::      :::   :::|||:::|||   |||:::   :
3666 TGCAAAAAATGTAACAGTAAACAACAATATTACTTCTCACAAAGCAGTGA 3715

  48 spLeuGly..................................... 50
     ::::::::
3716 GCATCTCTGCGACAAGTGGAGAAATTACCACTAAAACAGGTACAACCATT 3765

  51 .............................................Le 51
                                                   ||
3766 AACGCAACCACTGGTAACGTGGAGATAACCGCTCAAACAGGTAGTATCCT 3815

  51 uSerProThrLeuSerAlaProAlaProGlyValLeuThrGlnThrProG 68
     |:::   :::   |||:::::::::::::   :::||||||   |||   |
3816 AGGTGGAATTGAGTCCAGCTCTGGCTCTGTAACACTTACTGCAACCGAGG 3865

  68 ly......ThrIleThrSerPheLeuLysAlaSerIleGlnAsnThrAsp 82
     ||      ::::::::::::::      :::::::::      |||:::
3866 GCGCTCTTGCTGTAAGCAATATTTCGGGCAACACCGTTACTGTTACTGCA 3915

  83 MetAsnGlnAspLeuAsnAlaLeuAlaAsnAsnValThrThrLysAlaAs 99
        :::   :::|||:::::|||||||:::::::   |||:::|||:::::
3916 AATAGCGGTGCATTAACCACTTTGGCAGGCTCT...ACAATTAAAGGAAC 3962

  99 nGluValValGlnThrGlnLeuArgGluGlnGlnAlaGluValGlyLysP 116
     :|||   |||   |||           |||   :::::::::|||
3963 CGAGAGTGTAACCACT......TCAAGTCAATCAGGCGATATCGGC.... 4002

 116 hePheAspIleSerGlyMetSerSerSerAlaValAlaLeuLeuAlaAla 132
                :::   :::|||:::::::::|||::::::   |||:::
4003 ..........GGTACGATTTCTGGTGGCACAGTAGAGGTTAAAGCAACC 4041

 133 AlaAsnThrLeuMetLeuThrLeuAsnGlnAlaAspSerLysLeuSerGl 149
     ::::::         ||||||   :::|||:::::::||||||:::   ::
4042 GAAAGT.........TTAACC...ACTCAATCCAATTCAAAAATTAAAGC 4079

 149 yLysLeuSerLeuValSerPheAspAlaAlaLysThrThrAlaSerSerM 166
     ::::   :::             :::|||:::   |||:::|||::::::
4080 AACAACAGGC............GAGGCTAACGTAACAAGTGCAACAGGTA 4117

 166 etMetArgGluGlyMetAsnAlaLeuSerGlySerIleSerGlnSerAla 182
       :::         :::::::::   :::   :::::::::   :::|||
4118 CAATTGGTGGTACGATTTCCGGTAATACGGTAAATGTTACGGCAAACGCT 4167

 183 LeuGlnLeuGlyIleThrGlyValGlyAlaLysLeuGluTyrLysGlyLe 199
        :::|||   :::   |||   ||||||   ::::::
4168 GGCGATTTAACAGTT...GGGAATGGCGCAGAAATTAAT.........GC 4205

 199 uGlnAsnGluArgGlyAlaLeuLysHisAsnAlaAlaLysIleAspLysL 216
        :::      :::::::|||:::   :::::::::          ||||
4206 GACAGAAGGAGCTGCAACCTTAACTACATCATCGGGC.........AAAT 4246

 216 euThrThrGluSerHisSerIleLysAsnValLeuAsnGlyGlnAsnSer 232
     ||||||||||||:::   |||      ::::::   ::::::::::::
4247 TAACTACCGAAGCTAGTTCA......CACATTACTTCAGCCAAGGGTCAG 4290

 233 ValLysLeuGlyAlaGluGlyValAspSerLeuLysSerLeuAsnMetLy 249
     |||::::|||:::|||::::::   :::         |||:::|||
4291 GTAAATCTTTCAGCTCAGGATGGTAGCGTTGCAGGAAGTATTAATGCCGC 4340

 249 sLysThrGlyThrAspAlaThrLysAsnLeuAsnAspAlaThrLeuLysS 266
       ::::::      ::::::|||   :::|||      :::|||:::|||:
4341 CAATGTGACACTAAATACTACAGGCACTTTA......ACTACCGTGAAGG 4384

 266 erAsnAlaGlyThrSerAlaThrGluSerLeuGlyIleLysAspSerAsn 282
     ::::::::   ::::::::::::   :::|||   |||:::::   :::
4385 GTTCAAACATTAATGCAACCAGCGGTACCTTGGTTATTAACGCAAAAGAC 4434

 283 LysGlnIleSerProGluHisGlnAlaIleLeuSerLysArgLeuGluSe 299
        ::::::         :::   |||   |||::::::::::
4435 GCTGAGCTA.........AATGGCGCAGCATTGGGTAACCACACAGTGGT 4475

 299 rValGluSerAspIleArgLeuGluGlnAsnThrMetAspMetThrArgI 316
```

```
          :::::::::                 ::::::::::::   |||
4476 AAATGCAACCAACGCAAATGGCTCCGGCAGCGTAATCGCGACAACC.... 4521

 316 leAspAlaArgLysMetGlnMetThrGlyAspLeuIleMetLysAsnSer 332
           :::   ::::::::::::|||||||||||||||      |||:::
4522 .....TCAAGCAGAGTGAACATCACTGGGGATTTAATCACAATAAATGGA 4566

 333 ValThrVal 335
     :::::::::
4567 TTAAATATC 4575

seq_name: /cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-728-470-5

seq_documentation_block:
; Sequence 5, Application US/08728470
; Patent No. 5928651
;  GENERAL INFORMATION:
;    APPLICANT:  Barenkamp, Stephen J
;    TITLE OF INVENTION:  High Molecular Weight Surface Proteins
;    TITLE OF INVENTION:  of No. 5928651-Typeable Haemophilus
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd.
;      STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;      STREET:  Bldg. 1
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/728,470
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/302,832
;      FILING DATE:  16-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US PCT/US93/02166
;      FILING DATE:  16-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9205704.1
;      FILING DATE:  16-MAR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berkstresser, Jerry W
;      REGISTRATION NUMBER:  22,651
;      REFERENCE/DOCKET NUMBER:  1038-633
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 415-0813
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9171 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-728-470-5


alignment_scores:
            Quality:  102.50              Length:    353
              Ratio:  0.483                 Gaps:     15
 Percent Similarity:  60.057   Percent Identity: 19.547

alignment_block:
 US-09-068-804-6 x US-08-728-470-5   ..

 Align seg 1/1  to: US-08-728-470-5  from: 1  to: 9171
```

```
    ::::::::
3716 GCATCTCTGCGACAAGTGGAGAAATTACCACTAAAACAGGTACAACCATT 3765

  51 ..............................................Le  51
                                                   ||
3766 AACGCAACCACTGGTAACGTGGAGATAACCGCTCAAACAGGTAGTATCCT 3815

  51 uSerProThrLeuSerAlaProAlaProGlyValLeuThrGlnThrProG  68
     |:::   :::   |||:::::::::::::   :::|||||||   |||   |
3816 AGGTGGAATTGAGTCCAGCTCTGGCTCTGTAACACTTACTGCAACCGAGG 3865

  68 ly......ThrIleThrSerPheLeuLysAlaSerIleGlnAsnThrAsp  82
     ||      ::::::::::::::      :::::::::      |||:::
3866 GCGCTCTTGCTGTAAGCAATATTTCGGGCAACACCGTTACTGTTACTGCA 3915

  83 MetAsnGlnAspLeuAsnAlaLeuAlaAsnAsnValThrThrLysAlaAs  99
        :::   :::|||:::::::||||||::::::   |||:::|||:::::
3916 AATAGCGGTGCATTAACCACTTTGGCAGGCTCT...ACAATTAAAGGAAC 3962

  99 nGluValValGlnThrGlnLeuArgGluGlnGlnAlaGluValGlyLysP 116
     :|||   |||   |||          |||   :::::::::|||
3963 CGAGAGTGTAACCACT......TCAAGTCAATCAGGCGATATCGGC.... 4002

 116 hePheAspIleSerGlyMetSerSerSerAlaValAlaLeuLeuAlaAla 132
               :::   :::|||:::::::::|||::::::   |||:::
4003 ..........GGTACGATTTCTGGTGGCACAGTAGAGGTTAAAGCAACC 4041

 133 AlaAsnThrLeuMetLeuThrLeuAsnGlnAlaAspSerLysLeuSerGl 149
     ::::::         ||||||   :::|||::::::|||||:::   ::
4042 GAAAGT.........TTAACC...ACTCAATCCAATTCAAAAATTAAAGC 4079

 149 yLysLeuSerLeuValSerPheAspAlaAlaLysThrThrAlaSerSerM 166
     ::::   :::            :::|||:::   |||:::|||::::::
4080 AACAACAGGC............GAGGCTAACGTAACAAGTGCAACAGGTA 4117

 166 etMetArgGluGlyMetAsnAlaLeuSerGlySerIleSerGlnSerAla 182
       :::         :::::::::   :::   :::::::::   :::|||
4118 CAATTGGTGGTACGATTTCCGGTAATACGGTAAATGTTACGGCAAACGCT 4167

 183 LeuGlnLeuGlyIleThrGlyValGlyAlaLysLeuGluTyrLysGlyLe 199
        :::|||   :::   |||   ||||||  ::::::
4168 GGCGATTTAACAGTT...GGGAATGGCGCAGAAATTAAT.........GC 4205

 199 uGlnAsnGluArgGlyAlaLeuLysHisAsnAlaAlaLysIleAspLysL 216
        :::      ::::::|||:::   :::::::::         ||||
4206 GACAGAAGGAGCTGCAACCTTAACTACATCATCGGGC.........AAAT 4246

 216 euThrThrGluSerHisSerIleLysAsnValLeuAsnGlyGlnAsnSer 232
     ||||||||||::   |||      ::::::   ::::::::::::
4247 TAACTACCGAAGCTAGTTCA......CACATTACTTCAGCCAAGGGTCAG 4290

 233 ValLysLeuGlyAlaGluGlyValAspSerLeuLysSerLeuAsnMetLy 249
     |||:::|||:::|||:::::   :::         |||:::|||
4291 GTAAATCTTTCAGCTCAGGATGGTAGCGTTGCAGGAAGTATTAATGCCGC 4340

 249 sLysThrGlyThrAspAlaThrLysAsnLeuAsnAspAlaThrLeuLysS 266
      ::::::      ::::::|||   :::|||      :::|||:::|||:
4341 CAATGTGACACTAAATACTACAGGCACTTTA......ACTACCGTGAAGG 4384

 266 erAsnAlaGlyThrSerAlaThrGluSerLeuGlyIleLysAspSerAsn 282
     ::::::::   ::::::::::   :::|||   |||::::::   :::
4385 GTTCAAACATTAATGCAACCAGCGGTACCTTGGTTATTAACGCAAAAGAC 4434

 283 LysGlnIleSerProGluHisGlnAlaIleLeuSerLysArgLeuGluSe 299
     ::::::         :::   |||   |||::::::::::
4435 GCTGAGCTA.........AATGGCGCAGCATTGGGTAACCACACAGTGGT 4475

 299 rValGluSerAspIleArgLeuGluGlnAsnThrMetAspMetThrArgI 316
     :::::::::                  ::::::::::::   |||
4476 AAATGCAACCAACGCAAATGGCTCCGGCAGCGTAATCGCGACAACC.... 4521

 316 leAspAlaArgLysMetGlnMetThrGlyAspLeuIleMetLysAsnSer 332
       :::   ::::::::::::|||||||||||||||      |||:::
```

```
4522 .....TCAAGCAGAGTGAACATCACTGGGGATTTAATCACAATAAATGGA 4566

 333 ValThrVal 335
     :::::::::
4567 TTAAATATC 4575
```

seq_name: /cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-469-880-5

```
seq_documentation_block:
; Sequence 5, Application US/08469880
; Patent No. 5876733
;  GENERAL INFORMATION:
;    APPLICANT:  Barenkamp, Stephen J.
;    TITLE OF INVENTION:  High Molecular Weight Surface Proteins
;    TITLE OF INVENTION:  of No. 5876733-Typeable Haemophilus
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd.
;      STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;      STREET:  Bldg. 1
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/469,880
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9205704.1
;      FILING DATE:  16-MAR-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US PCT/US93/02166
;      FILING DATE:  16-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/302,832
;      FILING DATE:  16-SEP-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berkstresser, Jerry W
;      REGISTRATION NUMBER:  22,651
;      REFERENCE/DOCKET NUMBER:  1038-516 MIS:vg
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 415-0813
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9171 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-469-880-5
```

```
alignment_scores:
            Quality:  102.50              Length:    353
              Ratio:  0.483                 Gaps:     15
 Percent Similarity:  60.057   Percent Identity: 19.547

alignment_block:
 US-09-068-804-6 x US-08-469-880-5   ..

 Align seg 1/1  to: US-08-469-880-5  from: 1  to: 9171

      15 AsnAsnHisSerValGluAsnSerSerGlnThrAlaSerGlnSerValSe 31
         ||||||::::::   ::::::|||:::::::::|||:::   ::::::::
    3616 AATAACAACACTGAAGATAGCAGTGACAATAATGCCGGCTTAACTATCGA 3665
```

```
3666 TGCAAAAAATGTAACAGTAAACAACAATATTACTTCTCACAAAGCAGTGA 3715

  48 spLeuGly..................................... 50
     ::::::::
3716 GCATCTCTGCGACAAGTGGAGAAATTACCACTAAAACAGGTACAACCATT 3765

  51 .............................................Le 51
                                                   ||
3766 AACGCAACCACTGGTAACGTGGAGATAACCGCTCAAACAGGTAGTATCCT 3815

  51 uSerProThrLeuSerAlaProAlaProGlyValLeuThrGlnThrProG 68
     |:::   :::   |||:::::::::::::   :::|||||||   |||   |
3816 AGGTGGAATTGAGTCCAGCTCTGGCTCTGTAACACTTACTGCAACCGAGG 3865

  68 ly......ThrIleThrSerPheLeuLysAlaSerIleGlnAsnThrAsp 82
     ||      ::::::::::::::::      :::::::::      |||:::
3866 GCGCTCTTGCTGTAAGCAATATTTCGGGCAACACCGTTACTGTTACTGCA 3915

  83 MetAsnGlnAspLeuAsnAlaLeuAlaAsnAsnValThrThrLysAlaAs 99
        :::   :::|||:::::::||||||:::::::   |||:::|||:::::
3916 AATAGCGGTGCATTAACCACTTTGGCAGGCTCT...ACAATTAAAGGAAC 3962

  99 nGluValValGlnThrGlnLeuArgGluGlnGlnAlaGluValGlyLysP 116
     :|||   |||   |||            |||   ::::::::::|||
3963 CGAGAGTGTAACCACT......TCAAGTCAATCAGGCGATATCGGC.... 4002

 116 hePheAspIleSerGlyMetSerSerSerAlaValAlaLeuLeuAlaAla 132
               :::   :::|||::::::::::|||::::::   |||:::
4003 ...........GGTACGATTTCTGGTGGCACAGTAGAGGTTAAAGCAACC 4041

 133 AlaAsnThrLeuMetLeuThrLeuAsnGlnAlaAspSerLysLeuSerGl 149
     ::::::         ||||||   :::|||::::::||||||:::   ::
4042 GAAAGT.........TTAACC...ACTCAATCCAATTCAAAAATTAAAGC 4079

 149 yLysLeuSerLeuValSerPheAspAlaAlaLysThrThrAlaSerSerM 166
     ::::   :::            :::|||:::   |||:::|||::::::
4080 AACAACAGGC............GAGGCTAACGTAACAAGTGCAACAGGTA 4117

 166 etMetArgGluGlyMetAsnAlaLeuSerGlySerIleSerGlnSerAla 182
       :::         :::::::::   :::   :::::::::   :::|||
4118 CAATTGGTGGTACGATTTCCGGTAATACGGTAAATGTTACGGCAAACGCT 4167

 183 LeuGlnLeuGlyIleThrGlyValGlyAlaLysLeuGluTyrLysGlyLe 199
        :::|||   :::   |||   ||||||   ::::::
4168 GGCGATTTAACAGTT...GGGAATGGCGCAGAAATTAAT.........GC 4205

 199 uGlnAsnGluArgGlyAlaLeuLysHisAsnAlaAlaLysIleAspLysL 216
        :::       :::::::|||:::   :::::::::         ||||
4206 GACAGAAGGAGCTGCAACCTTAACTACATCATCGGGC.........AAAT 4246

 216 euThrThrGluSerHisSerIleLysAsnValLeuAsnGlyGlnAsnSer 232
     ||||||||||||:::   |||      ::::::   ::::::::::::
4247 TAACTACCGAAGCTAGTTCA......CACATTACTTCAGCCAAGGGTCAG 4290

 233 ValLysLeuGlyAlaGluGlyValAspSerLeuLysSerLeuAsnMetLy 249
     |||:::|||:::|||:::::::   :::         |||:::|||
4291 GTAAATCTTTCAGCTCAGGATGGTAGCGTTGCAGGAAGTATTAATGCCGC 4340

 249 sLysThrGlyThrAspAlaThrLysAsnLeuAsnAspAlaThrLeuLysS 266
       ::::::      ::::::|||   :::|||      :::|||:::|||:
4341 CAATGTGACACTAAATACTACAGGCACTTTA......ACTACCGTGAAGG 4384

 266 erAsnAlaGlyThrSerAlaThrGluSerLeuGlyIleLysAspSerAsn 282
     ::::::::   ::::::::::::   :::|||   |||::::::   :::
4385 GTTCAAACATTAATGCAACCAGCGGTACCTTGGTTATTAACGCAAAAGAC 4434

 283 LysGlnIleSerProGluHisGlnAlaIleLeuSerLysArgLeuGluSe 299
       ::::::         :::   |||   |||:::::::::
4435 GCTGAGCTA.........AATGGCGCAGCATTGGGTAACCACACAGTGGT 4475

 299 rValGluSerAspIleArgLeuGluGlnAsnThrMetAspMetThrArgI 316
        :::::::::          ::::::::::   |||
4476 AAATGCAACCAACGCAAATGGCTCCGGCAGCGTAATCGCGACAACC.... 4521
```

```
 316 leAspAlaArgLysMetGlnMetThrGlyAspLeuIleMetLysAsnSer 332
           :::   ::::::::::::|||||||||||||||      |||:::
4522 .....TCAAGCAGAGTGAACATCACTGGGGATTTAATCACAATAAATGGA 4566

 333 ValThrVal 335
     :::::::::
4567 TTAAATATC 4575
```

```
seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-530-198-5

seq_documentation_block:
; Sequence 5, Application US/08530198
; Patent No. 5869065
;  GENERAL INFORMATION:
;    APPLICANT:  BARENKAMP, STEPHEN J
;    APPLICANT:  ST. GEME III, JOSEPH W
;    TITLE OF INVENTION:  HIGH MOLECULAR WEIGHT SURFACE PROTEINS
;    TITLE OF INVENTION:  OF NON-TYPEABLE HAEMOPHILUS
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd
;      STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;      STREET:  Bldg. 1
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/530,198
;      FILING DATE:  13-DEC-1995
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BERKSTRESSER, JERRY W
;      REGISTRATION NUMBER:  22,651
;      REFERENCE/DOCKET NUMBER:  JWB-1186
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 415-0813
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9171 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-530-198-5


alignment_scores:
            Quality:  102.50              Length:    353
              Ratio:  0.483                 Gaps:     15
 Percent Similarity:  60.057   Percent Identity: 19.547

alignment_block:
 US-09-068-804-6 x US-08-530-198-5   ..

 Align seg 1/1  to: US-08-530-198-5  from: 1  to: 9171

      15 AsnAsnHisSerValGluAsnSerSerGlnThrAlaSerGlnSerValSe 31
         ||||||::::::   :::::::|||:::::::::::|||:::   ::::::::
    3616 AATAACAACACTGAAGATAGCAGTGACAATAATGCCGGCTTAACTATCGA 3665

      31 rAlaLysAspIleLeuAsnSerIleGlyIleSerSerSerLysValSerA 48
         :||||||::::::      :::   :::|||:::|||   |||:::   :
    3666 TGCAAAAAATGTAACAGTAAACAACAATATTACTTCTCACAAAGCAGTGA 3715

      48 spLeuGly..................................... 50
```

```
  48 spLeuGly..................................... 50
     ::::::::
3716 GCATCTCTGCGACAAGTGGAGAAATTACCACTAAAACAGGTACAACCATT 3765

  51 .............................................Le 51
                                                  ||
3766 AACGCAACCACTGGTAACGTGGAGATAACCGCTCAAACAGGTAGTATCCT 3815

  51 uSerProThrLeuSerAlaProAlaProGlyValLeuThrGlnThrProG 68
     |:::   :::   |||:::::::::::   :::||||||   |||   |
3816 AGGTGGAATTGAGTCCAGCTCTGGCTCTGTAACACTTACTGCAACCGAGG 3865

  68 ly......ThrIleThrSerPheLeuLysAlaSerIleGlnAsnThrAsp 82
     ||      ::::::::::::::      :::::::::      |||:::
3866 GCGCTCTTGCTGTAAGCAATATTTCGGGCAACACCGTTACTGTTACTGCA 3915

  83 MetAsnGlnAspLeuAsnAlaLeuAlaAsnAsnValThrThrLysAlaAs 99
        :::   :::|||:::::::||||||:::::::   |||:::|||:::::
3916 AATAGCGGTGCATTAACCACTTTGGCAGGCTCT...ACAATTAAAGGAAC 3962

  99 nGluValValGlnThrGlnLeuArgGluGlnGlnAlaGluValGlyLysP 116
     :|||   |||   |||            |||   :::::::::|||
3963 CGAGAGTGTAACCACT......TCAAGTCAATCAGGCGATATCGGC.... 4002

 116 hePheAspIleSerGlyMetSerSerSerAlaValAlaLeuLeuAlaAla 132
           :::   :::|||::::::::::|||::::::   |||:::
4003 ...........GGTACGATTTCTGGTGGCACAGTAGAGGTTAAAGCAACC 4041

 133 AlaAsnThrLeuMetLeuThrLeuAsnGlnAlaAspSerLysLeuSerGl 149
     ::::::         ||||||   :::|||:::::::|||||||::   ::
4042 GAAAGT.........TTAACC...ACTCAATCCAATTCAAAAATTAAAGC 4079

 149 yLysLeuSerLeuValSerPheAspAlaAlaLysThrThrAlaSerSerM 166
     ::::   :::            :::|||:::   |||:::|||::::::
4080 AACAACAGGC............GAGGCTAACGTAACAAGTGCAACAGGTA 4117

 166 etMetArgGluGlyMetAsnAlaLeuSerGlySerIleSerGlnSerAla 182
       :::         :::::::::   :::   :::::::::   :::|||
4118 CAATTGGTGGTACGATTTCCGGTAATACGGTAAATGTTACGGCAAACGCT 4167

 183 LeuGlnLeuGlyIleThrGlyValGlyAlaLysLeuGluTyrLysGlyLe 199
     :::|||   :::   |||   ||||||   ::::::
4168 GGCGATTTAACAGTT...GGGAATGGCGCAGAAATTAAT.........GC 4205

 199 uGlnAsnGluArgGlyAlaLeuLysHisAsnAlaAlaLysIleAspLysL 216
       :::       ::::::|||:::   :::::::::          ||||
4206 GACAGAAGGAGCTGCAACCTTAACTACATCATCGGGC.........AAAT 4246

 216 euThrThrGluSerHisSerIleLysAsnValLeuAsnGlyGlnAsnSer 232
     ||||||||||||:::   |||       ::::::   ::::::::::::
4247 TAACTACCGAAGCTAGTTCA......CACATTACTTCAGCCAAGGGTCAG 4290

 233 ValLysLeuGlyAlaGluGlyValAspSerLeuLysSerLeuAsnMetLy 249
     |||:::|||:::|||::::::   :::          |||:::|||
4291 GTAAATCTTTCAGCTCAGGATGGTAGCGTTGCAGGAAGTATTAATGCCGC 4340

 249 sLysThrGlyThrAspAlaThrLysAsnLeuAsnAspAlaThrLeuLysS 266
       ::::::       ::::::|||   :::|||       :::|||:::|||:
4341 CAATGTGACACTAAATACTACAGGCACTTTA......ACTACCGTGAAGG 4384

 266 erAsnAlaGlyThrSerAlaThrGluSerLeuGlyIleLysAspSerAsn 282
     ::::::::   ::::::::::::   :::|||   |||::::::   :::
4385 GTTCAAACATTAATGCAACCAGCGGTACCTTGGTTATTAACGCAAAAGAC 4434

 283 LysGlnIleSerProGluHisGlnAlaIleLeuSerLysArgLeuGluSe 299
       ::::::         :::   |||   |||::::::::::
4435 GCTGAGCTA.........AATGGCGCAGCATTGGGTAACCACACAGTGGT 4475

 299 rValGluSerAspIleArgLeuGluGlnAsnThrMetAspMetThrArgI 316
       :::::::::                ::::::::::   |||
4476 AAATGCAACCAACGCAAATGGCTCCGGCAGCGTAATCGCGACAACC.... 4521

 316 leAspAlaArgLysMetGlnMetThrGlyAspLeuIleMetLysAsnSer 332
```

```
                :::   ::::::::::::|||||||||||||||      |||:::
    4522 .....TCAAGCAGAGTGAACATCACTGGGGATTTAATCACAATAAATGGA 4566

     333 ValThrVal 335
         :::::::::
    4567 TTAAATATC 4575

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-302-832-5

seq_documentation_block:
; Sequence 5, Application US/08302832
; Patent No. 5603938
;  GENERAL INFORMATION:
;    APPLICANT:  Barenkamp, Stephen J
;    TITLE OF INVENTION:  High Molecular Weight Surface Proteins
;    TITLE OF INVENTION:  of No. 5603938-Typeable Haemophilus
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd.
;      STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;      STREET:  Bldg. 1
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/302,832
;      FILING DATE:  16-SEP-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9205704.1
;      FILING DATE:  16-MAR-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US pct/us93/02166
;      FILING DATE:  16-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berkstresser, Jerry W
;      REGISTRATION NUMBER:  22,651
;      REFERENCE/DOCKET NUMBER:  1038-404
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 415-0813
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9171 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-302-832-5


alignment_scores:
            Quality:  102.50              Length:    353
              Ratio:  0.483                 Gaps:     15
 Percent Similarity:  60.057   Percent Identity: 19.547

alignment_block:
 US-09-068-804-6 x US-08-302-832-5   ..

 Align seg 1/1  to: US-08-302-832-5  from: 1  to: 9171

      15 AsnAsnHisSerValGluAsnSerSerGlnThrAlaSerGlnSerValSe 31
         ||||||::::::   ::::::|||::::::::::|||:::   ::::::::
    3616 AATAACAACACTGAAGATAGCAGTGACAATAATGCCGGCTTAACTATCGA 3665

      31 rAlaLysAspIleLeuAsnSerIleGlyIleSerSerSerLysValSerA 48
         :||||||::::::      :::   :::|||:::|||   |||:::   :
```

```
    :::::      |||   ::::::|||||||
262 CGGGCAACGAGAAGCTAACCATGCAGAACCTCAACGACCGCCTGGCCTCC 311

 81 ...ThrAspMetAsnGlnAspLeuAsnAlaLeuAlaAsnAsnValThrTh  96
          |||:::   ::::::|||:::|||   ::::::::::::   ::
312 TACCTGGACAAGGTGCGCGCCCTGGAGGCGGCCAACGGCGAGCTAGAGGT 361

 96 rLysAlaAsnGluValValGlnThrGlnLeuArgGluGlnGlnAlaGluV 113
    :|||      :::      |||:::|||      :::      :::
362 GAAGATCCGCGACTGGTACCAGAAGCAGGGGCCTGGGCCCTCCCGCGACT 411

113 alGlyLysPhePhe...AspIleSerGlyMetSerSerSerAlaValAla 128
     :::   ::::::      |||   ::::::   :::      ::::::
412 ACAGCCACTACTACACGACCATCCAGGACCTGCGGGACAAGATTCTTGGT 461

129 LeuLeuAlaAlaAlaAsnThrLeuMetLeuThrLeuAsnGlnAlaAspSe 145
          |||:::   ::::::      ::::::|||:::   ::::::::
462 ......GCCACCATTGAGAACTCCAGGATTGTCCTGCAGATCGACAACGC 505

145 rLysLeuSerGlyLysLeuSerLeuValSerPheAspAlaAlaLysThr. 161
    ::::|||::::::          :::   |||:::::::::::::::::
506 CCGTCTGGCTGCAGATGACTTCCGAACCAAGTTTGAGACGGAACAGGCTC 555

162 .................ThrAlaSerSerMetMetArgGluGlyMetAsn 172
                    ::::::   :::   :::|||      ::::::
556 TGCGCATGAGCGTGGAGGCCGACATCAACGGCCTGCGCAGGGTGCTGGAT 605

173 AlaLeuSerGlySerIleSerGlnSerAlaLeuGlnLeuGlyIleThrGl 189
    :::|||      ::::::::::::::::::|||::::::   |||   ||
606 GAGCTG......ACCCTGGCCAGGACCGACCTGGAGATGCAGATCGAAGG 649

189 yValGlyAlaLysLeuGluTyrLysGlyLeuGlnAsnGluArgGlyAlaL 206
    |:::   :::   |||:::|||
650 CCTCAAGGAAGAGCTGGCCTAC.......................CTGA 675

206 euLysHisAsnAlaAlaLysIleAspLysLeuThrThrGluSerHisSer 222
      |||:::::::::::::   |||:::::::|||
676 AGAAGAACCATGAGGAGGAAATCAGTACGCTG.................. 707

223 IleLysAsnValLeuAsnGlyGlnAsnSerValLysLeuGlyAlaGluGl 239
       ::::::   ::::::|||||||   ||||||   ::::::::::::
708 ...AGGGGCCAAGTGGGAGGCCAGGTGAGTGTGGAGGTGGATTCCGCT.. 752

239 yValAspSerLeuLysSerLeuAsnMetLysLysThrGlyThrAspAlaT 256
                                        :::||||||||||   :
753 ..............................CCGGGCACCGATCTCG 768

256 hrLysAsnLeuAsnAspAlaThrLeuLysSerAsnAlaGlyThrSerAla 272
    ::|||   |||:::|||      ::::::|||:::      :::   |||
769 CCAAGATCCTGAGTGAC......ATGCGAAGCCAATATGAGGTGATGGCC 812

273 ThrGluSerLeuGlyIleLysAspSerAsnLysGlnIleSerProGluHi 289
                                       |||         :::
813 ............................GAGCAGAACCGGAAGGATGC 832

289 sGlnAlaIleLeuSerLysArgLeuGluSerValGluSerAspIleArgL 306
     :::|||   ::::::   |||   |||   ::::::   ::::::
833 TGAAGCCTGGTTCACCAGCCGGACTGAAGAATTGAACCGGGAGGTCGCTG 882

306 euGluGlnAsnThrMetAspMetThrArg.........IleAspAlaArg 319
      :::   :::   ::::::|||:::|||         :::|||   |||
883 GCCACACGGAGCAGCTCCAGATGAGCAGGTCCGAGGTTACTGACCTGCGG 932

320 Lys..................MetGlnMetThrGlyAspLeuIleMetLy 330
    :::                 :::::::::   ::::::|||   |||||
933 CGCACCCTTCAGGGTCTTGAGATTGAGCTGCAGTCACAGCTGAGCATGAA 982

330 sAsnSerValThrValGlyGlyIleAlaGlyAlaSerGlyGlnTyrAlaA 347
    |:::::::::      :::   :::|||   :::   ::::::::::::|
983 AGCTGCCTTGGAA...GACACACTGGCAGAAACGGAGGCGCGCTTTGGAG 1029

347 laThrGlnGluArgSerGluGlnGlnIleSerGlnValAsnAsnArgVal 363
    ||      ::::::   :::      ||||||   ::::::::::::::::
```

```
1030 CCCAGCTGGCGCATATCCAGGCGCTGATCAGCGGTATTGAAGCCCAGCTG 1079

 364 AlaSerThrAlaSerAspGluAlaArgGluSerSerArgLysSerThrSe 380
     |||::::::   :::|||   :::|||::::::
1080 GCGGATGTGCGAGCTGATAGTGAGCGGCAGAAT.................. 1112

 380 rLeuIleGlnGluMetLeuLysThrMetGlu 390
           ||||||      :::   |||:::
1113 .......CAGGAGTACCAGCGGCTCATGGAC 1136
```

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-038-682-5

```
seq_documentation_block:
; Sequence 5, Application US/08038682
; Patent No. 5549897
;  GENERAL INFORMATION:
;    APPLICANT:  BARENKAMP, STEPHEN J
;    APPLICANT:  ST. GEME III, JOSEPH W
;    TITLE OF INVENTION:  HIGH MOLECULAR WEIGHT SURFACE PROTEINS
;    TITLE OF INVENTION:  OF NON-TYPEABLE HAEMOPHILUS
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd
;      STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;      STREET:  Bldg. 1
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/038,682
;      FILING DATE:  16-MAR-1993
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BERKSTRESSER, JERRY W
;      REGISTRATION NUMBER:  22,651
;      REFERENCE/DOCKET NUMBER:  1038-293
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 415-0813
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9171 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-038-682-5


alignment_scores:
            Quality:  102.50            Length:    353
              Ratio:  0.483               Gaps:     15
 Percent Similarity:  60.057   Percent Identity: 19.547

alignment_block:
 US-09-068-804-6 x US-08-038-682-5   ..

 Align seg 1/1  to: US-08-038-682-5  from: 1  to: 9171

      15 AsnAsnHisSerValGluAsnSerSerGlnThrAlaSerGlnSerValSe 31
         ||||||::::::   :::::::|||:::::::::|||:::   ::::::::
    3616 AATAACAACACTGAAGATAGCAGTGACAATAATGCCGGCTTAACTATCGA 3665

      31 rAlaLysAspIleLeuAsnSerIleGlyIleSerSerSerLysValSerA 48
         :||||||::::::      :::   ::::|||:::|||   |||:::   :
    3666 TGCAAAAAATGTAACAGTAAACAACAATATTACTTCTCACAAAGCAGTGA 3715
```

```
  51 ..........................................Le  51
                                                ||
3816 AACGCAACCACTGGTAACGTGGAGATAACCGCTCAAACAGGTAGTATCCT 3865

  51 uSerProThrLeuSerAlaProAlaProGlyValLeuThrGlnThrProG  68
     |:::   :::   |||:::::::::::   :::|||||||   |||   |
3866 AGGTGGAATTGAGTCCAGCTCTGGCTCTGTAACACTTACTGCAACCGAGG 3915

  68 ly......ThrIleThrSerPheLeuLysAlaSerIleGlnAsnThrAsp  82
     ||      ::::::::::::::      :::::::::      |||:::
3916 GCGCTCTTGCTGTAAGCAATATTTCGGGCAACACCGTTACTGTTACTGCA 3965

  83 MetAsnGlnAspLeuAsnAlaLeuAlaAsnAsnValThrThrLysAlaAs  99
        :::   :::|||::::::|||||:::::   |||:::|||:::::
3966 AATAGCGGTGCATTAACCACTTTGGCAGGCTCT...ACAATTAAAGGAAC 4012

  99 nGluValValGlnThrGlnLeuArgGluGlnGlnAlaGluValGlyLysP 116
     :|||   |||   |||            |||   ::::::::::|||
4013 CGAGAGTGTAACCACT......TCAAGTCAATCAGGCGATATCGGC.... 4052

 116 hePheAspIleSerGlyMetSerSerSerAlaValAlaLeuLeuAlaAla 132
              :::   :::|||::::::::::|||:::::::   |||:::
4053 ...........GGTACGATTTCTGGTGGCACAGTAGAGGTTAAAGCAACC 4091

 133 AlaAsnThrLeuMetLeuThrLeuAsnGlnAlaAspSerLysLeuSerGl 149
     ::::::         ||||||   :::|||:::::::|||||||:::   ::
4092 GAAAGT.........TTAACC...ACTCAATCCAATTCAAAAATTAAAGC 4129

 149 yLysLeuSerLeuValSerPheAspAlaAlaLysThrThrAlaSerSerM 166
     ::::   :::            :::|||:::   |||:::|||::::::
4130 AACAACAGGC............GAGGCTAACGTAACAAGTGCAACAGGTA 4167

 166 etMetArgGluGlyMetAsnAlaLeuSerGlySerIleSerGlnSerAla 182
       :::         :::::::::   :::   :::::::::   :::|||
4168 CAATTGGTGGTACGATTTCCGGTAATACGGTAAATGTTACGGCAAACGCT 4217

 183 LeuGlnLeuGlyIleThrGlyValGlyAlaLysLeuGluTyrLysGlyLe 199
     :::|||   :::   |||   ||||||   ::::::
4218 GGCGATTTAACAGTT...GGGAATGGCGCAGAAATTAAT.........GC 4255

 199 uGlnAsnGluArgGlyAlaLeuLysHisAsnAlaAlaLysIleAspLysL 216
       :::       :::::::|||:::   :::::::::          ||||
4256 GACAGAAGGAGCTGCAACCTTAACTACATCATCGGGC.........AAAT 4296

 216 euThrThrGluSerHisSerIleLysAsnValLeuAsnGlyGlnAsnSer 232
     ||||||||||||:::   |||       ::::::   ::::::::::::
4297 TAACTACCGAAGCTAGTTCA......CACATTACTTCAGCCAAGGGTCAG 4340

 233 ValLysLeuGlyAlaGluGlyValAspSerLeuLysSerLeuAsnMetLy 249
     |||:::|||:::|||::::::   :::          |||:::|||
4341 GTAAATCTTTCAGCTCAGGATGGTAGCGTTGCAGGAAGTATTAATGCCGC 4390

 249 sLysThrGlyThrAspAlaThrLysAsnLeuAsnAspAlaThrLeuLysS 266
       ::::::      ::::::|||   :::|||       :::|||:::|||:
4391 CAATGTGACACTAAATACTACAGGCACTTTA......ACTACCGTGAAGG 4434

 266 erAsnAlaGlyThrSerAlaThrGluSerLeuGlyIleLysAspSerAsn 282
     ::::::::   ::::::::::::   :::|||   |||::::::   :::
4435 GTTCAAACATTAATGCAACCAGCGGTACCTTGGTTATTAACGCAAAAGAC 4484

 283 LysGlnIleSerProGluHisGlnAlaIleLeuSerLysArgLeuGluSe 299
     ::::::          :::   |||   |||::::::::::
4485 GCTGAGCTA.........AATGGCGCAGCATTGGGTAACCACACAGTGGT 4525

 299 rValGluSerAspIleArgLeuGluGlnAsnThrMetAspMetThrArgI 316
       :::::::::                 ::::::::::   |||
4526 AAATGCAACCAACGCAAATGGCTCCGGCAGCGTAATCGCGACAACC.... 4571

 316 leAspAlaArgLysMetGlnMetThrGlyAspLeuIleMetLysAsnSer 332
       :::   ::::::::::::|||||||||||||||      |||:::
4572 .....TCAAGCAGAGTGAACATCACTGGGGATTTAATCACAATAAATGGA 4616
```

```
      333 ValThrVal 335
          :::::::::
     4617 TTAAATATC 4625

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-730-953-1

seq_documentation_block:
; Sequence 1, Application US/07730953
; Patent No. 5288614
;  GENERAL INFORMATION:
;    APPLICANT:  BODENMULLER, Heinz
;    APPLICANT:  DESSAUER, Andreas
;    TITLE OF INVENTION:  METHOD FOR THE DETECTION OF MALIGNANT
;    TITLE OF INVENTION:  DISEASES
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nikaido, Marmelstein, Murray & Oram
;      STREET:  655 Fifteenth Street N.W. Suite 330
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-5701
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/730,953
;      FILING DATE:  19910723
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DE P 39 42 999.7
;      FILING DATE:  21-DEC-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kitts, Monica C.
;      REGISTRATION NUMBER:  36,105
;      REFERENCE/DOCKET NUMBER:  P564-1119
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)638-5000
;      TELEFAX:  (202)638-4810
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1394 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  33..1232
US-07-730-953-1


alignment_scores:
            Quality:   94.00              Length:    377
              Ratio:  0.459                Gaps:     14
 Percent Similarity:  54.377   Percent Identity: 17.241

alignment_block:
 US-09-068-804-6 x US-07-730-953-1   ..

 Align seg 1/1  to: US-07-730-953-1  from: 1  to: 1394

      38 SerIleGlyIleSerSerSerLysValSerAspLeuGlyLeuSerProTh 54
         ::::::::::::|||||::::::      :::   :::   |||
     162 GGCGTATCCGTGTCCTCCGCCCGCTTTGTGTCCTCGTCCTCCTCGGGGGG 211

      54 rLeuSerAlaProAlaProGlyValLeuThrGlnThrProGlyThrIleT 71
         .  ::::::      ||||||||||||   :::   |||   ::::
     212 CTACGGCGGCGGCTACGGCGGCGTCCTGACCGCGTCCGACGGGCTGCTGG 261

      71 hrSerPheLeuLysAlaSerIleGlnAsn..................... 80
```

```
  51 uSerProThrLeuSerAlaProAlaProGlyValLeuThrGlnThrProG 68
     |:::  :::  |||:::::::::::::  :::||||||  |||  |
3866 AGGTGGAATTGAGTCCAGCTCTGGCTCTGTAACACTTACTGCAACCGAGG 3915

  68 ly......ThrIleThrSerPheLeuLysAlaSerIleGlnAsnThrAsp 82
     ||      ::::::::::::::      :::::::::      |||:::
3916 GCGCTCTTGCTGTAAGCAATATTTCGGGCAACACCGTTACTGTTACTGCA 3965

  83 MetAsnGlnAspLeuAsnAlaLeuAlaAsnAsnValThrThrLysAlaAs 99
        :::   ::::|||::::::||||||::::::   |||:::|||:::::
3966 AATAGCGGTGCATTAACCACTTTGGCAGGCTCT...ACAATTAAAGGAAC 4012

  99 nGluValValGlnThrGlnLeuArgGluGlnGlnAlaGluValGlyLysP 116
     :|||  |||  |||           |||  :::::::::|||
4013 CGAGAGTGTAACCACT......TCAAGTCAATCAGGCGATATCGGC.... 4052

 116 hePheAspIleSerGlyMetSerSerSerAlaValAlaLeuLeuAlaAla 132
               :::  :::|||:::::::::|||::::::  |||:::
4053 ..........GGTACGATTTCTGGTGGCACAGTAGAGGTTAAAGCAACC 4091

 133 AlaAsnThrLeuMetLeuThrLeuAsnGlnAlaAspSerLysLeuSerGl 149
     ::::::         ||||||  :::|||:::::::||||||:::  ::
4092 GAAAGT.........TTAACC...ACTCAATCCAATTCAAAAATTAAAGC 4129

 149 yLysLeuSerLeuValSerPheAspAlaAlaLysThrThrAlaSerSerM 166
     ::::  :::            :::|||:::  |||:::|||::::::
4130 AACAACAGGC............GAGGCTAACGTAACAAGTGCAACAGGTA 4167

 166 etMetArgGluGlyMetAsnAlaLeuSerGlySerIleSerGlnSerAla 182
      :::        :::::::::  :::  :::::::::  :::|||
4168 CAATTGGTGGTACGATTTCCGGTAATACGGTAAATGTTACGGCAAACGCT 4217

 183 LeuGlnLeuGlyIleThrGlyValGlyAlaLysLeuGluTyrLysGlyLe 199
        :::|||  :::  |||  ||||||  ::::::
4218 GGCGATTTAACAGTT...GGGAATGGCGCAGAAATTAAT.........GC 4255

 199 uGlnAsnGluArgGlyAlaLeuLysHisAsnAlaAlaLysIleAspLysL 216
       :::      ::::::|||:::  :::::::::        ||||
4256 GACAGAAGGAGCTGCAACCTTAACTACATCATCGGGC.........AAAT 4296

 216 euThrThrGluSerHisSerIleLysAsnValLeuAsnGlyGlnAsnSer 232
     ||||||||||||:::  |||      ::::::  ::::::::::::
4297 TAACTACCGAAGCTAGTTCA......CACATTACTTCAGCCAAGGGTCAG 4340

 233 ValLysLeuGlyAlaGluGlyValAspSerLeuLysSerLeuAsnMetLy 249
     |||:::|||:::|||::::::  :::         |||:::|||
4341 GTAAATCTTTCAGCTCAGGATGGTAGCGTTGCAGGAAGTATTAATGCCGC 4390

 249 sLysThrGlyThrAspAlaThrLysAsnLeuAsnAspAlaThrLeuLysS 266
      ::::::     ::::::|||  :::|||      :::|||:::|||:
4391 CAATGTGACACTAAATACTACAGGCACTTTA......ACTACCGTGAAGG 4434

 266 erAsnAlaGlyThrSerAlaThrGluSerLeuGlyIleLysAspSerAsn 282
     ::::::::  ::::::::::::  :::|||  |||::::::  :::
4435 GTTCAAACATTAATGCAACCAGCGGTACCTTGGTTATTAACGCAAAAGAC 4484

 283 LysGlnIleSerProGluHisGlnAlaIleLeuSerLysArgLeuGluSe 299
        ::::::         :::  |||  |||::::::::::
4485 GCTGAGCTA.........AATGGCGCAGCATTGGGTAACCACACAGTGGT 4525

 299 rValGluSerAspIleArgLeuGluGlnAsnThrMetAspMetThrArgI 316
        :::::::::              ::::::::::  |||
4526 AAATGCAACCAACGCAAATGGCTCCGGCAGCGTAATCGCGCACAACC.... 4571

 316 leAspAlaArgLysMetGlnMetThrGlyAspLeuIleMetLysAsnSer 332
       :::  ::::::::::::|||||||||||||||  |||:::
4572 .....TCAAGCAGAGTGAACATCACTGGGGATTTAATCACAATAAATGGA 4616

 333 ValThrVal 335
     :::::::::
4617 TTAAATATC 4625

seq_name: /cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-728-470-1
```

```
seq_documentation_block:
; Sequence 1, Application US/08728470
; Patent No. 5928651
;  GENERAL INFORMATION:
;    APPLICANT:  Barenkamp, Stephen J
;    TITLE OF INVENTION:  High Molecular Weight Surface Proteins
;    TITLE OF INVENTION:  of No. 5928651-Typeable Haemophilus
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd.
;      STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;      STREET:  Bldg. 1
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/728,470
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/302,832
;      FILING DATE:  16-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US PCT/US93/02166
;      FILING DATE:  16-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9205704.1
;      FILING DATE:  16-MAR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berkstresser, Jerry W
;      REGISTRATION NUMBER:  22,651
;      REFERENCE/DOCKET NUMBER:  1038-633
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 415-0813
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5116 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-728-470-1


alignment_scores:
            Quality:  102.50              Length:    353
              Ratio:  0.483                 Gaps:     15
 Percent Similarity:  60.057   Percent Identity: 19.547

alignment_block:
 US-09-068-804-6 x US-08-728-470-1   ..

 Align seg 1/1  to: US-08-728-470-1  from: 1  to: 5116

      15 AsnAsnHisSerValGluAsnSerSerGlnThrAlaSerGlnSerValSe 31
         ||||||::::::   ::::::|||::::::::::||||:::   :::::::
    3666 AATAACAACACTGAAGATAGCAGTGACAATAATGCCGGCTTAACTATCGA 3715

      31 rAlaLysAspIleLeuAsnSerIleGlyIleSerSerSerLysValSerA 48
         :||||||::::::       :::   ::::|||:::|||   |||:::   :
    3716 TGCAAAAAATGTAACAGTAAACAACAATATTACTTCTCACAAAGCAGTGA 3765

      48 spLeuGly..................................... 50
         ::::::::
    3766 GCATCTCTGCGACAAGTGGAGAAATTACCACTAAAACAGGTACAACCATT 3815
```

```
     ||      ::::::::::::::::      :::::::::      |||:::
3916 GCGCTCTTGCTGTAAGCAATATTTCGGGCAACACCGTTACTGTTACTGCA 3965

  83 MetAsnGlnAspLeuAsnAlaLeuAlaAsnAsnValThrThrLysAlaAs 99
     :::   ::::|||::::::||||||:::::::   |||:::|||:::::
3966 AATAGCGGTGCATTAACCACTTTGGCAGGCTCT...ACAATTAAAGGAAC 4012

  99 nGluValValGlnThrGlnLeuArgGluGlnGlnAlaGluValGlyLysP 116
     :|||   |||   |||             |||   ::::::::::|||
4013 CGAGAGTGTAACCACT......TCAAGTCAATCAGGCGATATCGGC.... 4052

 116 hePheAspIleSerGlyMetSerSerSerAlaValAlaLeuLeuAlaAla 132
               :::   :::|||:::::::::|||::::::   |||:::
4053 ..........GGTACGATTTCTGGTGGCACAGTAGAGGTTAAAGCAACC 4091

 133 AlaAsnThrLeuMetLeuThrLeuAsnGlnAlaAspSerLysLeuSerGl 149
     ::::::          ||||||   :::|||::::::||||||:::   ::
4092 GAAAGT.........TTAACC...ACTCAATCCAATTCAAAAATTAAAGC 4129

 149 yLysLeuSerLeuValSerPheAspAlaAlaLysThrThrAlaSerSerM 166
     ::::   :::             :::|||:::   |||:::|||::::::
4130 AACAACAGGC............GAGGCTAACGTAACAAGTGCAACAGGTA 4167

 166 etMetArgGluGlyMetAsnAlaLeuSerGlySerIleSerGlnSerAla 182
       :::         :::::::::   :::   :::::::::   :::|||
4168 CAATTGGTGGTACGATTTCCGGTAATACGGTAAATGTTACGGCAAACGCT 4217

 183 LeuGlnLeuGlyIleThrGlyValGlyAlaLysLeuGluTyrLysGlyLe 199
       :::|||   :::   |||   ||||||   ::::::
4218 GGCGATTTAACAGTT...GGGAATGGCGCAGAAATTAAT.........GC 4255

 199 uGlnAsnGluArgGlyAlaLeuLysHisAsnAlaAlaLysIleAspLysL 216
       :::      :::::::|||:::   :::::::::         ||||
4256 GACAGAAGGAGCTGCAACCTTAACTACATCATCGGGC.........AAAT 4296

 216 euThrThrGluSerHisSerIleLysAsnValLeuAsnGlyGlnAsnSer 232
     ||||||||||||:::   |||      ::::::   ::::::::::::
4297 TAACTACCGAAGCTAGTTCA......CACATTACTTCAGCCAAGGGTCAG 4340

 233 ValLysLeuGlyAlaGluGlyValAspSerLeuLysSerLeuAsnMetLy 249
     |||:::|||:::|||::::::   :::         |||:::|||
4341 GTAAATCTTTCAGCTCAGGATGGTAGCGTTGCAGGAAGTATTAATGCCGC 4390

 249 sLysThrGlyThrAspAlaThrLysAsnLeuAsnAspAlaThrLeuLysS 266
      ::::::      ::::::|||   :::|||      :::|||:::|||:
4391 CAATGTGACACTAAATACTACAGGCACTTTA......ACTACCGTGAAGG 4434

 266 erAsnAlaGlyThrSerAlaThrGluSerLeuGlyIleLysAspSerAsn 282
     ::::::::   ::::::::::::   :::|||   |||::::::   :::
4435 GTTCAAACATTAATGCAACCAGCGGTACCTTGGTTATTAACGCAAAAGAC 4484

 283 LysGlnIleSerProGluHisGlnAlaIleLeuSerLysArgLeuGluSe 299
       ::::::          :::   |||   |||:::::::::
4485 GCTGAGCTA.........AATGGCGCAGCATTGGGTAACCACACAGTGGT 4525

 299 rValGluSerAspIleArgLeuGluGlnAsnThrMetAspMetThrArgI 316
       :::::::::                 ::::::::::::   |||
4526 AAATGCAACCAACGCAAATGGCTCCGGCAGCGTAATCGCGACAACC.... 4571

 316 leAspAlaArgLysMetGlnMetThrGlyAspLeuIleMetLysAsnSer 332
       :::   ::::::::::::|||||||||||||||      |||:::
4572 .....TCAAGCAGAGTGAACATCACTGGGGATTTAATCACAATAAATGGA 4616

 333 ValThrVal 335
     :::::::::
4617 TTAAATATC 4625

seq_name: /cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-469-880-1

seq_documentation_block:
; Sequence 1, Application US/08469880
; Patent No. 5876733
;  GENERAL INFORMATION:
;    APPLICANT:  Barenkamp, Stephen J.
;    TITLE OF INVENTION:  High Molecular Weight Surface Proteins
;    TITLE OF INVENTION:  of No. 5876733-Typeable Haemophilus
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd.
;      STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;      STREET:  Bldg. 1
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/469,880
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9205704.1
;      FILING DATE:  16-MAR-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US PCT/US93/02166
;      FILING DATE:  16-MAR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/302,832
;      FILING DATE:  16-SEP-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berkstresser, Jerry W
;      REGISTRATION NUMBER:  22,651
;      REFERENCE/DOCKET NUMBER:  1038-516 MIS:vg
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 415-0813
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5116 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-469-880-1


alignment_scores:
            Quality:  102.50              Length:    353
              Ratio:  0.483                Gaps:     15
 Percent Similarity:  60.057   Percent Identity: 19.547

alignment_block:
 US-09-068-804-6 x US-08-469-880-1   ..

 Align seg 1/1  to: US-09-068-804-6  from: 1  to: 5116

      15 AsnAsnHisSerValGluAsnSerSerGlnThrAlaSerGlnSerValSe 31
         ||||||::::::   ::::::|||:::::::::|||:::   ::::::::
    3666 AATAACAACACTGAAGATAGCAGTGACAATAATGCCGGCTTAACTATCGA 3715

      31 rAlaLysAspIleLeuAsnSerIleGlyIleSerSerSerLysValSerA 48
         :||||||::::::      :::   :::|||:::|||   ||:::   :
    3716 TGCAAAAAATGTAACAGTAAACAACAATATTACTTCTCACAAAGCAGTGA 3765

      48 spLeuGly..................................... 50
         ::::::::
    3766 GCATCTCTGCGACAAGTGGAGAAATTACCACTAAAACAGGTACAACCATT 3815

      51 .....................................------Le 51
                                                     ||
    3816 AACGCAACCACTGGTAACGTGGAGATAACCGCTCAAACAGGTAGTATCCT 3865
```

```
3866 AGGTGGAATTGAGTCCAGCTCTGGCTCTGTAACACTTACTGCAACCGAGG 3915

  68 ly......ThrIleThrSerPheLeuLysAlaSerIleGlnAsnThrAsp 82
     ||      ::::::::::::::      :::::::::       |||:::
3916 GCGCTCTTGCTGTAAGCAATATTTCGGGCAACACCGTTACTGTTACTGCA 3965

  83 MetAsnGlnAspLeuAsnAlaLeuAlaAsnAsnValThrThrLysAlaAs 99
        :::   :::|||:::::::||||||:::::::   |||:::|||:::::
3966 AATAGCGGTGCATTAACCACTTTGGCAGGCTCT...ACAATTAAAGGAAC 4012

  99 nGluValValGlnThrGlnLeuArgGluGlnGlnAlaGluValGlyLysP 116
     :|||   |||   |||             |||   :::::::::|||
4013 CGAGAGTGTAACCACT......TCAAGTCAATCAGGCGATATCGGC.... 4052

 116 hePheAspIleSerGlyMetSerSerSerAlaValAlaLeuLeuAlaAla 132
               :::   :::|||::::::::|||:::::::   |||:::
4053 ...........GGTACGATTTCTGGTGGCACAGTAGAGGTTAAAGCAACC 4091

 133 AlaAsnThrLeuMetLeuThrLeuAsnGlnAlaAspSerLysLeuSerGl 149
     ::::::         ||||||   :::|||:::::::||||||:::   ::
4092 GAAAGT.........TTAACC...ACTCAATCCAATTCAAAAATTAAAGC 4129

 149 yLysLeuSerLeuValSerPheAspAlaAlaLysThrThrAlaSerSerM 166
     ::::   :::            :::|||:::   |||:::|||:::::::
4130 AACAACAGGC...........GAGGCTAACGTAACAAGTGCAACAGGTA 4167

 166 etMetArgGluGlyMetAsnAlaLeuSerGlySerIleSerGlnSerAla 182
       :::         :::::::::   :::   :::::::::   :::|||
4168 CAATTGGTGGTACGATTTCCGGTAATACGGTAAATGTTACGGCAAACGCT 4217

 183 LeuGlnLeuGlyIleThrGlyValGlyAlaLysLeuGluTyrLysGlyLe 199
       :::|||   :::   |||   ||||||   ::::::
4218 GGCGATTTAACAGTT...GGGAATGGCGCAGAAATTAAT.........GC 4255

 199 uGlnAsnGluArgGlyAlaLeuLysHisAsnAlaAlaLysIleAspLysL 216
       :::      ::::::|||:::   :::::::::         ||||
4256 GACAGAAGGAGCTGCAACCTTAACTACATCATCGGGC.........AAAT 4296

 216 euThrThrGluSerHisSerIleLysAsnValLeuAsnGlyGlnAsnSer 232
     ||||||||||:::   |||      ::::::   ::::::::::::
4297 TAACTACCGAAGCTAGTTCA......CACATTACTTCAGCCAAGGGTCAG 4340

 233 ValLysLeuGlyAlaGluGlyValAspSerLeuLysSerLeuAsnMetLy 249
     |||:::|||:::|||:::::::   :::       |||:::|||
4341 GTAAATCTTTCAGCTCAGGATGGTAGCGTTGCAGGAAGTATTAATGCCGC 4390

 249 sLysThrGlyThrAspAlaThrLysAsnLeuAsnAspAlaThrLeuLysS 266
      ::::::      ::::::|||   :::|||      :::|||:::|||:
4391 CAATGTGACACTAAATACTACAGGCACTTTA......ACTACCGTGAAGG 4434

 266 erAsnAlaGlyThrSerAlaThrGluSerLeuGlyIleLysAspSerAsn 282
     ::::::::   ::::::::::::   :::|||   |||::::::   :::
4435 GTTCAAACATTAATGCAACCAGCGGTACCTTGGTTATTAACGCAAAAGAC 4484

 283 LysGlnIleSerProGluHisGlnAlaIleLeuSerLysArgLeuGluSe 299
       ::::::         :::   |||   |||::::::::::
4485 GCTGAGCTA.........AATGGCGCAGCATTGGGTAACCACACAGTGGT 4525

 299 rValGluSerAspIleArgLeuGluGlnAsnThrMetAspMetThrArgI 316
       :::::::::                 ::::::::::   |||
4526 AAATGCAACCAACGCAAATGGCTCCGGCAGCGTAATCGCGACAACC.... 4571

 316 leAspAlaArgLysMetGlnMetThrGlyAspLeuIleMetLysAsnSer 332
           :::   ::::::::::::|||||||||||||||      |||:::
4572 .....TCAAGCAGAGTGAACATCACTGGGGATTTAATCACAATAAATGGA 4616

 333 ValThrVal 335
     :::::::::
4617 TTAAATATC 4625

seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-530-198-1

seq_documentation_block:

; Sequence 1, Application US/08530198
; Patent No. 5869065
; GENERAL INFORMATION:
;   APPLICANT: BARENKAMP, STEPHEN J
;   APPLICANT: ST. GEME III, JOSEPH W
;   TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
;   TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
;   NUMBER OF SEQUENCES: 8
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Shoemaker and Mattare, Ltd
;     STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;     STREET: Bldg. 1
;     CITY: Arlington
;     STATE: Virginia
;     COUNTRY: U.S.A.
;     ZIP: 22202-0286
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/530,198
;     FILING DATE: 13-DEC-1995
;     CLASSIFICATION: 424
;   ATTORNEY/AGENT INFORMATION:
;     NAME: BERKSTRESSER, JERRY W
;     REGISTRATION NUMBER: 22,651
;     REFERENCE/DOCKET NUMBER: JWB-1186
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (703) 415-0810
;     TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 5116 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
US-08-530-198-1


alignment_scores:
            Quality:  102.50              Length:    353
              Ratio:  0.483                 Gaps:     15
 Percent Similarity:  60.057   Percent Identity: 19.547

alignment_block:
 US-09-068-804-6 x US-08-530-198-1   ..

 Align seg 1/1  to: US-08-530-198-1  from: 1  to: 5116

      15 AsnAsnHisSerValGluAsnSerSerGlnThrAlaSerGlnSerValSe 31
         ||||||::::::   :::::::|||:::::::::::|||:::   ::::::::
    3666 AATAACAACACTGAAGATAGCAGTGACAATAATGCCGGCTTAACTATCGA 3715

      31 rAlaLysAspIleLeuAsnSerIleGlyIleSerSerSerLysValSerA 48
         :||||||::::::       :::   :::|||:::|||   |||:::   :
    3716 TGCAAAAAATGTAACAGTAAACAACAATATTACTTCTCACAAAGCAGTGA 3765

      48 spLeuGly..................................... 50
         ::::::::
    3766 GCATCTCTGCGACAAGTGGAGAAATTACCACTAAAACAGGTACAACCATT 3815

      51 .....................................................Le 51
                                                             ||
    3816 AACGCAACCACTGGTAACGTGGAGATAACCGCTCAAACAGGTAGTATCCT 3865

      51 uSerProThrLeuSerAlaProAlaProGlyValLeuThrGlnThrProG 68
         |:::   :::   |||:::::::::::::   :::||||||   |||   |
    3866 AGGTGGAATTGAGTCCAGCTCTGGCTCTGTAACACTTACTGCAACCGAGG 3915

      68 ly......ThrIleThrSerPheLeuLysAlaSerIleGlnAsnThrAsp 82
```

```
     68 ly.....ThrIleThrSerPheLeuLysAlaSerIleGlnAsnThrAsp 82
        ||      :::::::::::::::      :::::::::      |||:::
   3916 GCGCTCTTGCTGTAAGCAATATTTCGGGCAACACCGTTACTGTTACTGCA 3965

     83 MetAsnGlnAspLeuAsnAlaLeuAlaAsnAsnValThrThrLysAlaAs 99
        :::   :::|||::::::|||||:::::   |||:::|||:::::
   3966 AATAGCGGTGCATTAACCACTTTGGCAGGCTCT...ACAATTAAAGGAAC 4012

     99 nGluValValGlnThrGlnLeuArgGluGlnGlnAlaGluValGlyLysP 116
        :|||  |||  |||           |||  :::::::::|||
   4013 CGAGAGTGTAACCACT......TCAAGTCAATCAGGCGATATCGGC.... 4052

    116 hePheAspIleSerGlyMetSerSerSerAlaValAlaLeuLeuAlaAla 132
              :::  :::|||:::::::::|||::::::  |||:::
   4053 ..........GGTACGATTTCTGGTGGCACAGTAGAGGTTAAAGCAACC 4091

    133 AlaAsnThrLeuMetLeuThrLeuAsnGlnAlaAspSerLysLeuSerGl 149
        ::::::         ||||||  :::|||::::::||||||:::  ::
   4092 GAAAGT.....TTAACC...ACTCAATCCAATTCAAAAATTAAAGC 4129

    149 yLysLeuSerLeuValSerPheAspAlaAlaLysThrThrAlaSerSerM 166
        ::::  :::           :::|||:::  |||:::|||:::::
   4130 AACAACAGGC..........GAGGCTAACGTAACAAGTGCAACAGGTA 4167

    166 etMetArgGluGlyMetAsnAlaLeuSerGlySerIleSerGlnSerAla 182
          :::        :::::::::  :::  :::::::::  :::|||
   4168 CAATTGGTGGTACGATTTCCGGTAATACGGTAAATGTTACGGCAAACGCT 4217

    183 LeuGlnLeuGlyIleThrGlyValGlyAlaLysLeuGluTyrLysGlyLe 199
        :::|||  :::  |||  ||||||  ::::::
   4218 GGCGATTTAACAGTT...GGGAATGGCGCAGAAATTAAT.....GC 4255

    199 uGlnAsnGluArgGlyAlaLeuLysHisAsnAlaAlaLysIleAspLysL 216
        :::      ::::::|||:::  :::::::::       ||||
   4256 GACAGAAGGAGCTGCAACCTTAACTACATCATCGGGC.....AAAT 4296

    216 euThrThrGluSerHisSerIleLysAsnValLeuAsnGlyGlnAsnSer 232
        ||||||||||:::  |||     ::::::  ::::::::::::
   4297 TAACTACCGAAGCTAGTTCA......CACATTACTTCAGCCAAGGGTCAG 4340

    233 ValLysLeuGlyAlaGluGlyValAspSerLeuLysSerLeuAsnMetLy 249
        |||:::|||:::|||:::::::  :::        |||:::|||
   4341 GTAAATCTTTCAGCTCAGGATGGTAGCGTTGCAGGAAGTATTAATGCCGC 4390

    249 sLysThrGlyThrAspAlaThrLysAsnLeuAsnAspAlaThrLeuLysS 266
        ::::::     ::::::|||  :::|||      :::|||:::|||:
   4391 CAATGTGACACTAAATACTACAGGCACTTTA......ACTACCGTGAAGG 4434

    266 erAsnAlaGlyThrSerAlaThrGluSerLeuGlyIleLysAspSerAsn 282
        ::::::::  ::::::::::::  :::|||  |||::::::  :::
   4435 GTTCAAACATTAATGCAACCAGCGGTACCTTGGTTATTAACGCAAAAGAC 4484

    283 LysGlnIleSerProGluHisGlnAlaIleLeuSerLysArgLeuGluSe 299
        ::::::        :::  |||  |||:::::::::
   4485 GCTGAGCTA.....AATGGCGCAGCATTGGGTAACCACACAGTGGT 4525

    299 rValGluSerAspIleArgLeuGluGlnAsnThrMetAspMetThrArgI 316
        :::::::::          ::::::::::  |||
   4526 AAATGCAACCAACGCAAATGGCTCCGGCAGCGTAATCGCGACAACC.... 4571

    316 leAspAlaArgLysMetGlnMetThrGlyAspLeuIleMetLysAsnSer 332
          :::  ::::::::::::|||||||||||||||     |||:::
   4572 .....TCAAGCAGAGTGAACATCACTGGGGATTTAATCACAATAAATGGA 4616

    333 ValThrVal 335
        :::::::::
   4617 TTAAATATC 4625

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-302-832-1

seq_documentation_block:
; Sequence 1, Application US/08302832
; Patent No. 5603938

;  GENERAL INFORMATION:
;    APPLICANT:  Barenkamp, Stephen J
;    TITLE OF INVENTION:  High Molecular Weight Surface Proteins
;    TITLE OF INVENTION:  of No. 5603938-Typeable Haemophilus
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd.
;      STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;      STREET:  Bldg. 1
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/302,832
;      FILING DATE:  16-SEP-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9205704.1
;      FILING DATE:  16-MAR-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US pct/us93/02166
;      FILING DATE:  16-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berkstresser, Jerry W
;      REGISTRATION NUMBER:  22,651
;      REFERENCE/DOCKET NUMBER:  1038-404
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 415-0813
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5116 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-302-832-1


alignment_scores:
            Quality:  102.50              Length:    353
              Ratio:  0.483                 Gaps:     15
 Percent Similarity:  60.057   Percent Identity: 19.547

alignment_block:
 US-09-068-804-6 x US-08-302-832-1   ..

 Align seg 1/1  to: US-08-302-832-1  from: 1  to: 5116

     15 AsnAsnHisSerValGluAsnSerSerGlnThrAlaSerGlnSerValSe 31
        ||||||:::::::   ::::::|||::::::::::|||:::   ::::::::
   3666 AATAACAACACTGAAGATAGCAGTGACAATAATGCCGGCTTAACTATCGA 3715

     31 rAlaLysAspIleLeuAsnSerIleGlyIleSerSerSerLysValSerA 48
        :||||||:::::: .    :::   :::|||:::|||   |||:::   :
   3716 TGCAAAAAATGTAACAGTAAACAACAATATTACTTCTCACAAAGCAGTGA 3765

     48 spLeuGly..................................... 50
        ::::::::
   3766 GCATCTCTGCGACAAGTGGAGAAATTACCACTAAAACAGGTACAACCATT 3815

     51 ..........................................Le 51
                                                   ||
   3816 AACGCAACCACTGGTAACGTGGAGATAACCGCTCAAACAGGTAGTATCCT 3865

     51 uSerProThrLeuSerAlaProAlaProGlyValLeuThrGlnThrProG 68
        |:::   :::   |||:::::::::::::   :::||||||   |||   |
```

```
2143 GGACGCATTGCAGACGCACCAGAGAAAGCTGCAGGAAGCGCTCGAGAGCT 2192

 186 lyIleThrGlyValGlyAlaLys.....LeuGluTyrLysGly 198
2193 CTGAGCGGACAGCCGCGGAAAGGGACCAGCTGCTGCAGCAGCTAACAGAG 2242

 199 LeuGlnAsnGluArgGlyAlaLeuLysHisAsnAlaAlaLysIleAspLy 215
2243 CTTCAGTCTGAGCGTACGCAGCTATCACAGGTTGTGACCGACCGCGAGCG 2292

 215 sLeuThrThrGluSerHisSerIleLys..... 224
2293 GCTTACACGCGACTTGCAGCGTATTCAGTACGAGTACGGGGAAACCGAGC 2342

 224 ..... 224
2343 TCGCGCGAGACGTGGCGCTGTGCGCCGCGCAGGAGATGGAGGCGCGCTAC 2392

 225 ......AsnValLeuAsnGlyGlnAsnSerValLysLeuGlyAlaGluGl 239
2393 CACGCTGCTGTGTTTCACCTGCAAACGCTCCTGGAGCTCGCAACCGAGTG 2442

 239 yValAspSerLeuLysSerLeuAsnMetLysLysThrGlyThrAspAlaT 256
2443 GGAGGACGCACTCCGCGAGCGTGCGCTTGCAGAGCGTGACGAAGCCGCTG 2492

 256 hrLysAsnLeuAsnAspAlaThrLeuLysSer...AsnAlaGlyThrSer 271
2493 CAGCCGAACTTGATGCCGCAGCCTCTACTTCCCAAAACGCACGTGAAAGC 2542

 272 AlaThrGluSerLeuGlyIleLysAspSerAsnLysGlnIleSerProGl 288
2543 GCCTGCGAGCGGCTAACCAGCCTTGAGCAGCAGCTTCGCGAATCCGAGGA 2592

 288 uHisGlnAlaIleLeuSerLysArgLeuGluSerValGluSerAspIleA 305
2593 GCGCGCTGCGGAGCTGGCGAGCCAGCTGGAGGCCACTGCTGCTGCGAAGT 2642

 305 rgLeuGluGlnAsnThrMetAspMetThrArgIle..... 316
2643 CGTCGGCGGAGCAGGACCGCGAGAACACGAGGGCCACGCTAGAGCAGCAG 2692

 317 .....AspAlaArgLysMetGlnMetThrGlyAspLeuIleMe 329
2693 CTTCGCGAATCCGAGGCGCGCGCTGCGGAGCTGGCGAGCCAGCTG..... 2737

 329 tLysAsnSerValThrValGlyGlyIleAlaGlyAlaSerGlyGlnTyrA 346
2738 .....GAGGCCACTGCTGCTGCGAAGATGT 2762

 346 laAlaThrGlnGluArgSer.....GluGlnGlnIle 356
2763 CAGCGGAGCAGGACCGCGAGAACACGAGGGCCACGCTAGAGCAGCAGCTT 2812

 357 SerGlnValAsnAsnArgValAlaSerThrAlaSer..... 368
2813 CGTGACTCCGAGGAGCGCGCTGCGGAGCTGGCGAGCCAGCTGGAGTCCAC 2862

 369 .....AspGluAlaArgGluSerSerArgLysS 378
2863 TACTGCTGCGAAGATGTCAGCGGAGCAGGACCGCGAGAGCACGAGG.... 2908

 378 erThrSerLeuIleGlnGluMetLeuLysThrMetGluSerIleAsnGln 394
2909 ..GCCACGCTAGAGCAGCAGCTTCGTGACTCCGAGGAGCGCGCTGCGGAG 2956

 395 SerLysAlaSerAlaLeuAlaAlaIleAlaGlyAsnIleArgAla*** 410
2957 CTGGCGAGCCAGCTGGAGTCCACTACTGCTGCGAAGATGTCAGCGGAG 3004

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-038-682-1
```

```
seq_documentation_block:
; Sequence 1, Application US/08038682
; Patent No. 5549897
;  GENERAL INFORMATION:
;    APPLICANT:  BARENKAMP, STEPHEN J
;    APPLICANT:  ST. GEME III, JOSEPH W
;    TITLE OF INVENTION:  HIGH MOLECULAR WEIGHT SURFACE PROTEINS
;    TITLE OF INVENTION:  OF NON-TYPEABLE HAEMOPHILUS
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd
;      STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;      STREET:  Bldg. 1
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/038,682
;      FILING DATE:  16-MAR-1993
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BERKSTRESSER, JERRY W
;      REGISTRATION NUMBER:  22,651
;      REFERENCE/DOCKET NUMBER:  1038-293
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 415-0813
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5116 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-038-682-1


alignment_scores:
            Quality:  102.50              Length:    353
              Ratio:  0.483                Gaps:     15
 Percent Similarity:  60.057   Percent Identity: 19.547

alignment_block:
 US-09-068-804-6 x US-08-038-682-1   ..

 Align seg 1/1  to: US-08-038-682-1  from: 1  to: 5116

      15 AsnAsnHisSerValGluAsnSerSerGlnThrAlaSerGlnSerValSe 31
    3666 AATAACAACACTGAAGATAGCAGTGACAATAATGCCGGCTTAACTATCGA 3715

      31 rAlaLysAspIleLeuAsnSerIleGlyIleSerSerSerLysValSerA 48
    3716 TGCAAAAAATGTAACAGTAAACAACAATATTACTTCTCACAAAGCAGTGA 3765

      48 spLeuGly..... 50
    3766 GCATCTCTGCGACAAGTGGAGAAATTACCACTAAAACAGGTACAACCATT 3815

      51 .....Le 51
    3816 AACGCAACCACTGGTAACGTGGAGATAACCGCTCAAACAGGTAGTATCCT 3865

      51 uSerProThrLeuSerAlaProAlaProGlyValLeuThrGlnThrProG 68
    3866 AGGTGGAATTGAGTCCAGCTCTGGCTCTGTAACACTTACTGCAACCGAGG 3915
```

```
     288 uHisGlnAlaIleLeuSerLysArgLeuGluSerValGluSerAspIleA 305
         |:::   |||   |||:::   :::||||||:::::::::::::::
    2593 GCGCGCTGCGGAGCTGGCGAGCCAGCTGGAGGCCACTGCTGCTGCGAAGT 2642

     305 rgLeuGluGlnAsnThrMetAspMetThrArgIle..............  316
           :::::::::      :::   ||||||
    2643 CGTCGGCGGAGCAGGACCGCGAGAACACGAGGGCCACGCTAGAGCAGCAG 2692

     317 ............AspAlaArgLysMetGlnMetThrGlyAspLeuIleMe 329
                   :::||||||      :::::::::::::::|||
    2693 CTTCGCGAATCCGAGGCGCGCGCTGCGGAGCTGGCGAGCCAGCTG..... 2737

     329 tLysAsnSerValThrValGlyGlyIleAlaGlyAlaSerGlyGlnTyrA 346
                                 |||   |||:::::::::   :
    2738 .......................GAGGCCACTGCTGCTGCGAAGATGT 2762

     346 laAlaThrGlnGluArgSer.................GluGlnGlnIle 356
         ::|||   |||:::|||                    |||||||||:::
    2763 CAGCGGAGCAGGACCGCGAGAACACGAGGGCCACGCTAGAGCAGCAGCTT 2812

     357 SerGlnValAsnAsnArgValAlaSerThrAlaSer.............. 368
           :::   ::::::|||:::|||      ||||||
    2813 CGTGACTCCGAGGAGCGCGCTGCGGAGCTGGCGAGCCAGCTGGAGTCCAC 2862

     369 .....................AspGluAlaArgGluSerSerArgLysS 378
                             :::::::::|||||||||:::|||
    2863 TACTGCTGCGAAGATGTCAGCGGAGCAGGACCGCGAGAGCACGAGG.... 2908

     378 erThrSerLeuIleGlnGluMetLeuLysThrMetGluSerIleAsnGln 394
           ::::::|||   |||:::::      :::   |||      :::::
    2909 ..GCCACGCTAGAGCAGCAGCTTCGTGACTCCGAGGAGCGCGCTGCGGAG 2956

     395 SerLysAlaSerAlaLeuAlaAlaIleAlaGlyAsnIleArgAla*** 410
           :::          :::::::::|||:::::::::   ||||||
    2957 CTGGCGAGCCAGCTGGAGTCCACTACTGCTGCGAAGATGTCAGCGGAG 3004
```

```
seq_name: /cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:PCT-US94-00324-2

seq_documentation_block:
; Sequence 2, Application PC/TUS9400324
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven
;    TITLE OF INVENTION:  Diagnosis of Leishmaniasis
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Immunex Corporation
;      STREET:  51 University Street
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98101
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Apple Macintosh
;      OPERATING SYSTEM:  Apple System 7.1
;      SOFTWARE:  Microsoft Word, version 5.1a
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/00324
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/006,676
;      FILING DATE:  15-JAN-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Perkins, Patricia Anne
;      REGISTRATION NUMBER:  34,693
;      REFERENCE/DOCKET NUMBER:  5004-WO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 587-0430
;      TELEFAX:  (206) 233-0644
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3319 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Leishmania chagasi
PCT-US94-00324-2


alignment_scores:
           Quality:  109.50              Length:    516
             Ratio:  0.429                 Gaps:     22
 Percent Similarity:  49.419   Percent Identity: 19.574

alignment_block:
 US-09-068-804-6 x PCT-US94-00324-2   ..

 Align seg 1/1  to: PCT-US94-00324-2  from: 1  to: 3319

       2 LeuIleSerAsnValGlyIleAsnProAlaAlaTyrLeuAsnAsnHisSe 18
         :::|||:::::::|||:::   ::::::   :::|||
    1547 ATGATCGCGACTGTGAGCCCGAGCGCGCTGAACTAC.............. 1582

      18 rValGluAsnSerSerGlnThrAlaSerGlnSerValSerAlaLysAspI 35
           |||::::::      |||          :::      |||:::||||
    1583 ....GAGGAGACGCTGAGCACGCTGCGGTACGCGTCGCGCGCGCGCACA 1628

      35 leLeuAsnSerIleGlyIleSer............SerSerLysValSer 47
         ||:::|||         ::::::            :::   ::::::
    1629 TTGTGAATGTTGCGCAGGTGAACGAGGACCCGCGCGCACGGCGGATCCGC 1678

      48 AspLeu..............GlyLeuSerProThrLeuSerAlaProAl 59
         :::|||              ::::::   ::::::::::::::::   ::
    1679 GAGCTGGAGGAGCAGATGGAGGACATGCGGCAGGCGATGGCTGGCGGCGA 1728

      59 aProGlyValLeuThrGlnThrProGlyThrIleThrSerPheLeuLysA 76
         :|||:::   :::::::::         ::::::   :::::|||   :
    1729 CCCCGCGTACGTGTCTGAGCTGAAGAAGAAGCTT...GCGCTGCTGGAGT 1775

      76 laSerIleGlnAsnThrAspMetAsnGlnAspLeuAsnAlaLeuAlaAsn 92
         ::      |||:::          :::   ||||||:::|||||||:::
    1776 CGGAGGCGCAGAAG......CGTGCGGCGGACCTGCAGGCGCTGGAGAGG 1819

      93 AsnValThrThrLysAlaAsnGluValValGlnThrGln...LeuArgGl 108
         :::         ::::::::::::   ||||||   :::   ||||||::
    1820 GAG.........CGGGAGCACAACCAGGTGCAGGAGCGGCTGCTGCGCGC 1860

     108 uGlnGlnAlaGluValGlyLysPhePheAspIleSerGlyMetSerSerS 125
         :   :::||||||                  |||   :::   |||
    1861 GACGGAGGCGGAG..................AAGAGCGAGCTGGAGTCGC 1892

     125 erAlaValAlaLeu.........LeuAlaAlaAla.............. 133
           |||:::||||||         :::::::|||:::
    1893 GTGCGGCTGCGCTGCAGGAGGAGATGACCGCGACTCGACGGCAGGCGGAC 1942

     134 ......AsnThrLeuMetLeuThrLeuAsnGlnAlaAspSerLysLeuSe 148
               :::::::|||   |||   |||::::::::::::::::::
    1943 AAGATGCAGGCGCTGAACCTGCGGCTGAAGGAAGAGCAGGCGCGCAAGGA 1992

     148 rGlyLysLeu......SerLeuValSerPheAspAlaAla.......... 159
           |||         ::::::      |||||||||
    1993 GCGCGAGCTGCTGAAAGAGATGGCGAAGAAGGACGCCGCGCTCTCGAAGG 2042

     160 .............................................Lys 160

    2043 TTCGGCGACGCAAAGACGCCGAGATAGCAAGCGAGCGCGAGAAGCTGGAG 2092

     161 ThrThrAlaSerSerMetMetArgGlu..................GlyMe 171
         :::|||::::::   :::   ||||||                :::::
    2093 TCGACCGTGGCGCAGCTGGAGCGTGAGCAGCGCGAGCGCGAGGTGGCTCT 2142

     171 tAsnAlaLeuSerGlySerIleSer......GlnSerAlaLeuGlnLeuG 186
```

```
    395 SerLysAlaSerAlaLeuAlaAlaIleAlaGlyAsnIleArgAla*** 410
           :::         :::::::::|||:::::::::   ||||||
   2957 CTGGCGAGCCAGCTGGAGTCCACTACTGCTGCGAAGATGTCAGCGGAG 3004

seq_name: /cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-428-414A-4

seq_documentation_block:
; Sequence 4, Application US/08428414A
; Patent No. 5912166
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven G.
;    TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR DIAGNOSIS OF
;    TITLE OF INVENTION:  LEISHMANIASIS
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/428,414A
;      FILING DATE:  21-APR-1995
;      CLASSIFICATION:  436
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kadlecek, Ann T.
;      REGISTRATION NUMBER:  39,244
;      REFERENCE/DOCKET NUMBER:  210121.407
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;      TELEX:  3723836 SEEDANDBERRY
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3319 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
US-08-428-414A-4


alignment_scores:
            Quality:  109.50              Length:    516
              Ratio:  0.429                Gaps:     22
 Percent Similarity:  49.419   Percent Identity: 19.574

alignment_block:
 US-09-068-804-6 x US-08-428-414A-4   ..

 Align seq 1/1  to: US-08-428-414A-4  from: 1  to: 3319

       2 LeuIleSerAsnValGlyIleAsnProAlaAlaTyrLeuAsnAsnHisSe 18
         :::|||:::::::|||:::   ::::::   :::|||
    1547 ATGATCGCGACTGTGAGCCCGAGCGCGCTGAACTAC.............. 1582

      18 rValGluAsnSerSerGlnThrAlaSerGlnSerValSerAlaLysAspI 35
           |||::::::      |||        :::      |||:::||||
    1583 ....GAGGAGACGCTGAGCACGCTGCGGTACGCGTCGCGCGCGCGCGACA 1628

      35 leLeuAsnSerIleGlyIleSer............SerSerLysValSer 47
         ||:::|||        ::::::           :::   ::::::
    1629 TTGTGAATGTTGCGCAGGTGAACGAGGACCCGCGCGCACGGCGGATCCGC 1678

      48 AspLeu..............GlyLeuSerProThrLeuSerAlaProAl 59
         :::|||              ::::::   ::::::::::::::::   ::

    1679 GAGCTGGAGGAGCAGATGGAGGACATGCGGCAGGCGATGGCTGGCGGCGA 1728

      59 aProGlyValLeuThrGlnThrProGlyThrIleThrSerPheLeuLysA 76
         :|||:::   :::::::::         ::::::   ::::::|||   :
    1729 CCCCGCGTACGTGTCTGAGCTGAAGAAGAAGCTT...GCGCTGCTGGAGT 1775

      76 laSerIleGlnAsnThrAspMetAsnGlnAspLeuAsnAlaLeuAlaAsn 92
         ::      |||:::          :::   ||||||:::|||||:::
    1776 CGGAGGCGCAGAAG......CGTGCGGCGGACCTGCAGGCGCTGGAGAGG 1819

      93 AsnValThrThrLysAlaAsnGluValValGlnThrGln...LeuArgGl 108
         :::         ::::::::::::   ||||||   :::   |||||||::
    1820 GAG.........CGGGAGCACAACCAGGTGCAGGAGCGGCTGCTGCGCGC 1860

     108 uGlnGlnAlaGluValGlyLysPhePheAspIleSerGlyMetSerSerS 125
         :   :::|||||||                        |||   :::   |||
    1861 GACGGAGGCGGAG.................AAGAGCGAGCTGGAGTCGC 1892

     125 erAlaValAlaLeu.........LeuAlaAlaAla.............. 133
           |||:::||||||         ::::::|||:::
    1893 GTGCGGCTGCGCTGCAGGAGGAGATGACCGCGACTCGACGGCAGGCGGAC 1942

     134 ......AsnThrLeuMetLeuThrLeuAsnGlnAlaAspSerLysLeuSe 148
               :::::::|||   |||   |||::::::::::::::::::::
    1943 AAGATGCAGGCGCTGAACCTGCGGCTGAAGGAAGAGCAGGCGCGCAAGGA 1992

     148 rGlyLysLeu......SerLeuValSerPheAspAlaAla.......... 159
               |||         ::::::      |||||||||
    1993 GCGCGAGCTGCTGAAAGAGATGGCGAAGAAGGACGCCGCGCTCTCGAAGG 2042

     160 ..................................................Lys 160

    2043 TTCGGCGACGCAAAGACGCCGAGATAGCAAGCGAGCGCGAGAAGCTGGAG 2092

     161 ThrThrAlaSerSerMetMetArgGlu..................GlyMe 171
         :::|||::::::   :::   ||||||                :::::
    2093 TCGACCGTGGCGCAGCTGGAGCGTGAGCAGCGCGAGCGCGAGGTGGCTCT 2142

     171 tAsnAlaLeuSerGlySerIleSer......GlnSerAlaLeuGlnLeuG 186
         :::::||||||                    |||   ||||||:::   :
    2143 GGACGCATTGCAGACGCACCAGAGAAAGCTGCAGGAAGCGCTCGAGAGCT 2192

     186 lyIleThrGlyValGlyAlaLys............LeuGluTyrLysGly 198
         ::          ::::::::::::                :::   :::
    2193 CTGAGCGGACAGCCGCGGAAAGGGACCAGCTGCTGCAGCAGCTAACAGAG 2242

     199 LeuGlnAsnGluArgGlyAlaLeuLysHisAsnAlaAlaLysIleAspLy 215
         ||||||:::||||||      |||   :::   ::::::      :::::
    2243 CTTCAGTCTGAGCGTACGCAGCTATCACAGGTTGTGACCGACCGCGAGCG 2292

     215 sLeuThrThrGluSerHisSerIleLys..................... 224
         :||||||   :::   :::   |||:::
    2293 GCTTACACGCGACTTGCAGCGTATTCAGTACGAGTACGGGGAAACCGAGC 2342

     224 ..................................................... 224

    2343 TCGCGCGAGACGTGGCGCTGTGCGCCGCGCAGGAGATGGAGGCGCGCTAC 2392

     225 ......AsnValLeuAsnGlyGlnAsnSerValLysLeuGlyAlaGluGl 239
               :::|||::::::   |||:::   :::   |||::::::|||
    2393 CACGCTGCTGTGTTTCACCTGCAAACGCTCCTGGAGCTCGCAACCGAGTG 2442

     239 yValAspSerLeuLysSerLeuAsnMetLysLysThrGlyThrAspAlaT 256
           |||:::|||:::      ::::::          :::   :::|||:
    2443 GGAGGACGCACTCCGCGAGCGTGCGCTTGCAGAGCGTGACGAAGCCGCTG 2492

     256 hrLysAsnLeuAsnAspAlaThrLeuLysSer...AsnAlaGlyThrSer 271
         ::   :::|||::::::|||:::   :::|||   ||||||      |||
    2493 CAGCCGAACTTGATGCCGCAGCCTCTACTTCCCAAAACGCACGTGAAAGC 2542

     272 AlaThrGluSerLeuGlyIleLysAspSerAsnLysGlnIleSerProGl 288
         |||   |||   |||          :::   :::   :::   |||   ||
    2543 GCCTGCGAGCGGCTAACCAGCCTTGAGCAGCAGCTTCGCGAATCCGAGGA 2592
```

```
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3319 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    IMMEDIATE SOURCE:
;      CLONE:  rK39
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  455..3319
US-08-282-845-1


alignment_scores:
            Quality:  109.50              Length:    516
              Ratio:  0.429                Gaps:     22
 Percent Similarity:  49.419   Percent Identity: 19.574

alignment_block:
 US-09-068-804-6 x US-08-282-845-1   ..

 Align seg 1/1  to: US-08-282-845-1  from: 1  to: 3319

       2 LeuIleSerAsnValGlyIleAsnProAlaAlaTyrLeuAsnAsnHisSe 18
         :::|||:::::::|||:::   ::::::   :::|||
    1547 ATGATCGCGACTGTGAGCCCGAGCGCGCTGAACTAC..............  1582

      18 rValGluAsnSerSerGlnThrAlaSerGlnSerValSerAlaLysAspI 35
           |||::::::      |||       :::     |||:::||||
    1583 ....GAGGAGACGCTGAGCACGCTGCGGTACGCGTCGCGCGCGCGCACA 1628

      35 leLeuAsnSerIleGlyIleSer............SerSerLysValSer 47
         ||:::|||        ::::::            :::   ::::::
    1629 TTGTGAATGTTGCGCAGGTGAACGAGGACCCGCGCGCACGGCGGATCCGC 1678

      48 AspLeu..............GlyLeuSerProThrLeuSerAlaProAl 59
         :::|||               ::::::   ::::::::::::::::   ::
    1679 GAGCTGGAGGAGCAGATGGAGGACATGCGGCAGGCGATGGCTGGCGGCGA 1728

      59 aProGlyValLeuThrGlnThrProGlyThrIleThrSerPheLeuLysA 76
         :|||:::   :::::::::          ::::::   ::::::|||   :
    1729 CCCCGCGTACGTGTCTGAGCTGAAGAAGAAGCTT...GCGCTGCTGGAGT 1775

      76 laSerIleGlnAsnThrAspMetAsnGlnAspLeuAsnAlaLeuAlaAsn 92
         ::      |||:::         :::   ||||||:::|||||:::
    1776 CGGAGGCGCAGAAG......CGTGCGGCGGACCTGCAGGCGCTGGAGAGG 1819

      93 AsnValThrThrLysAlaAsnGluValValGlnThrGln...LeuArgGl 108
         :::          ::::::::::::   ||||||   :::   ||||||::
    1820 GAG.........CGGGAGCACAACCAGGTGCAGGAGCGGCTGCTGCGCGC 1860

     108 uGlnGlnAlaGluValGlyLysPhePheAspIleSerGlyMetSerSerS 125
         :   :::|||||||                        |||   :::   |||
    1861 GACGGAGGCGGAG.....................AAGAGCGAGCTGGAGTCGC 1892

     125 erAlaValAlaLeu.........LeuAlaAlaAla.............. 133
           |||:::|||||          ::::::|||:::
    1893 GTGCGGCTGCGCTGCAGGAGGAGATGACCGCGACTCGACGGCAGGCGGAC 1942

     134 ......AsnThrLeuMetLeuThrLeuAsnGlnAlaAspSerLysLeuSe 148
               :::::::\\\   \\\   |||:::::::::::::::::::
    1943 AAGATGCAGGCGCTGAACCTGCGGCTGAAGGAAGAGCAGGCGCGCAAGGA 1992

     148 rGlyLysLeu......SerLeuValSerPheAspAlaAla.......... 159
               |||          ::::::      |||||||||
    1993 GCGCGAGCTGCTGAAAGAGATGGCGAAGAAGGACGCCGCGCTCTCGAAGG 2042

     160 ..................................................Lys 160

    2043 TTCGGCGACGCAAAGACGCCGAGATAGCAAGCGAGCGCGAGAAGCTGGAG 2092

     161 ThrThrAlaSerSerMetMetArgGlu.................GlyMe 171
         :::|||::::::   :::   ||||||                 :::::
    2093 TCGACCGTGGCGCAGCTGGAGCGTGAGCAGCGCGAGCGCGAGGTGGCTCT 2142

     171 tAsnAlaLeuSerGlySerIleSer......GlnSerAlaLeuGlnLeuG 186
         ::::||||||                    |||   ||||||:::   :
    2143 GGACGCATTGCAGACGCACCAGAGAAAGCTGCAGGAAGCGCTCGAGAGCT 2192

     186 lyIleThrGlyValGlyAlaLys............LeuGluTyrLysGly 198
         ::          ::::::::::::             :::   :::
    2193 CTGAGCGGACAGCCGCGGAAAGGGACCAGCTGCTGCAGCAGCTAACAGAG 2242

     199 LeuGlnAsnGluArgGlyAlaLeuLysHisAsnAlaAlaLysIleAspLy 215
         ||||||:::||||||      |||   :::   ::::::      :::::
    2243 CTTCAGTCTGAGCGTACGCAGCTATCACAGGTTGTGACCGACCGCGAGCG 2292

     215 sLeuThrThrGluSerHisSerIleLys..................... 224
         :||||||   :::   :::   |||:::
    2293 GCTTACACGCGACTTGCAGCGTATTCAGTACGAGTACGGGGAAACCGAGC 2342

     224 ..................................................... 224

    2343 TCGCGCGAGACGTGGCGCTGTGCGCCGCGCAGGAGATGGAGGCGCGCTAC 2392

     225 ......AsnValLeuAsnGlyGlnAsnSerValLysLeuGlyAlaGluGl 239
               :::|||::::::   |||:::   :::   |||::::::|||
    2393 CACGCTGCTGTGTTTCACCTGCAAACGCTCCTGGAGCTCGCAACCGAGTG 2442

     239 yValAspSerLeuLysSerLeuAsnMetLysLysThrGlyThrAspAlaT 256
         |||:::|||:::      ::::::          :::   :::|||:
    2443 GGAGGACGCACTCCGCGAGCGTGCGCTTGCAGAGCGTGACGAAGCCGCTG 2492

     256 hrLysAsnLeuAsnAspAlaThrLeuLysSer...AsnAlaGlyThrSer 271
         ::   :::|||::::::|||:::   :::|||   ||||||      |||
    2493 CAGCCGAACTTGATGCCGCAGCCTCTACTTCCCAAAACGCACGTGAAAGC 2542

     272 AlaThrGluSerLeuGlyIleLysAspSerAsnLysGlnIleSerProGl 288
         |||   |||   |||         :::   :::   :::   |||   ||
    2543 GCCTGCGAGCGGCTAACCAGCCTTGAGCAGCAGCTTCGCGAATCCGAGGA 2592

     288 uHisGlnAlaIleLeuSerLysArgLeuGluSerValGluSerAspIleA 305
         |:::   |||   |||:::   :::||||||:::::::::::::::
    2593 GCGCGCTGCGGAGCTGGCGAGCCAGCTGGAGGCCACTGCTGCTGCGAAGT 2642

     305 rgLeuGluGlnAsnThrMetAspMetThrArgIle.............. 316
         :::::::::       :::   ||||||
    2643 CGTCGGCGGAGCAGGACCGCGAGAACACGAGGGCCACGCTAGAGCAGCAG 2692

     317 ............AspAlaArgLysMetGlnMetThrGlyAspLeuIleMe 329
                     :::|||||||       :::::::::::::::|||
    2693 CTTCGCGAATCCGAGGCGCGCGCTGCGGAGCTGGCGAGCCAGCTG..... 2737

     329 tLysAsnSerValThrValGlyGlyIleAlaGlyAlaSerGlyGlnTyrA 346
                                       |||   |||:::::::::  :
    2738 .......................GAGGCCACTGCTGCTGCGAAGATGT 2762

     346 laAlaThrGlnGluArgSer.................GluGlnGlnIle 356
         ::|||   |||:::|||                   ||||||||||:::
    2763 CAGCGGAGCAGGACCGCGAGAACACGAGGGCCACGCTAGAGCAGCAGCTT 2812

     357 SerGlnValAsnAsnArgValAlaSerThrAlaSer.............. 368
           :::   ::::::|||:::|||       ||||||
    2813 CGTGACTCCGAGGAGCGCGCTGCGGAGCTGGCGAGCCAGCTGGAGTCCAC 2862

     369 .....................AspGluAlaArgGluSerSerArgLysS 378
                              :::::::::|||||||||:::|||
    2863 TACTGCTGCGAAGATGTCAGCGGAGCAGGACCGCGAGAGCACGAGG.... 2908

     378 erThrSerLeuIleGlnGluMetLeuLysThrMetGluSerIleAsnGln 394
           :::::::|||   |||:::::::      :::   |||      ::::::
    2909 ..GCCACGCTAGAGCAGCAGCTTCGTGACTCCGAGGAGCGCGCTGCGGAG 2956
```

```
 59 aProGlyValLeuThrGlnThrProGlyThrIleThrSerPheLeuLysA 76
1729 CCCCGCGTACGTGTCTGAGCTGAAGAAGAAGCTT...GCGCTGCTGGAGT 1775

 76 laSerIleGlnAsnThrAspMetAsnGlnAspLeuAsnAlaLeuAlaAsn 92
1776 CGGAGGCGCAGAAG......CGTGCGGCGGACCTGCAGGCGCTGGAGAGG 1819

 93 AsnValThrThrLysAlaAsnGluValValGlnThrGln...LeuArgGl 108
1820 GAG.........CGGGAGCACAACCAGGTGCAGGAGCGGCTGCTGCGCGC 1860

108 uGlnGlnAlaGluValGlyLysPhePheAspIleSerGlyMetSerSerS 125
1861 GACGGAGGCGGAG..................AAGAGCGAGCTGGAGTCGC 1892

125 erAlaValAlaLeu.........LeuAlaAlaAla.............. 133
1893 GTGCGGCTGCGCTGCAGGAGGAGATGACCGCGACTCGACGGCAGGCGGAC 1942

134 ......AsnThrLeuMetLeuThrLeuAsnGlnAlaAspSerLysLeuSe 148
1943 AAGATGCAGGCGCTGAACCTGCGGCTGAAGGAAGAGCAGGCGCGCAAGGA 1992

148 rGlyLysLeu......SerLeuValSerPheAspAlaAla.......... 159
1993 GCGCGAGCTGCTGAAAGAGATGGCGAAGAAGGACGCCGCGCTCTCGAAGG 2042

160 ..............................................Lys 160
2043 TTCGGCGACGCAAAGACGCCGAGATAGCAAGCGAGCGCGAGAAGCTGGAG 2092

161 ThrThrAlaSerSerMetMetArgGlu..................GlyMe 171
2093 TCGACCGTGGCGCAGCTGGAGCGTGAGCAGCGCGAGCGCGAGGTGGCTCT 2142

171 tAsnAlaLeuSerGlySerIleSer......GlnSerAlaLeuGlnLeuG 186
2143 GGACGCATTGCAGACGCACCAGAGAAAGCTGCAGGAAGCGCTCGAGAGCT 2192

186 lyIleThrGlyValGlyAlaLys............LeuGluTyrLysGly 198
2193 CTGAGCGGACAGCCGCGGAAAGGGACCAGCTGCTGCAGCAGCTAACAGAG 2242

199 LeuGlnAsnGluArgGlyAlaLeuLysHisAsnAlaAlaLysIleAspLy 215
2243 CTTCAGTCTGAGCGTACGCAGCTATCACAGGTTGTGACCGACCGCGAGCG 2292

215 sLeuThrThrGluSerHisSerIleLys..................... 224
2293 GCTTACACGCGACTTGCAGCGTATTCAGTACGAGTACGGGGAAACCGAGC 2342

224 ..............................................  224
2343 TCGCGCGAGACGTGGCGCTGTGCGCCGCGCAGGAGATGGAGGCGCGCTAC 2392

225 ......AsnValLeuAsnGlyGlnAsnSerValLysLeuGlyAlaGluGl 239
2393 CACGCTGCTGTGTTTCACCTGCAAACGCTCCTGGAGCTCGCAACCGAGTG 2442

239 yValAspSerLeuLysSerLeuAsnMetLysLysThrGlyThrAspAlaT 256
2443 GGAGGACGCACTCCGCGAGCGTGCGCTTGCAGAGCGTGACGAAGCCGCTG 2492

256 hrLysAsnLeuAsnAspAlaThrLeuLysSer...AsnAlaGlyThrSer 271
2493 CAGCCGAACTTGATGCCGCAGCCTCTACTTCCCAAAACGCACGTGAAAGC 2542

272 AlaThrGluSerLeuGlyIleLysAspSerAsnLysGlnIleSerProGl 288
2543 GCCTGCGAGCGGCTAACCAGCCTTGAGCAGCAGCTTCGCGAATCCGAGGA 2592

288 uHisGlnAlaIleLeuSerLysArgLeuGluSerValGluSerAspIleA 305
2593 GCGCGCTGCGGAGCTGGCGAGCCAGCTGGAGGCCACTGCTGCTGCGAAGT 2642

305 rgLeuGluGlnAsnThrMetAspMetThrArgIle.............. 316
2643 CGTCGGCGGAGCAGGACCGCGAGAACACGAGGGCCACGCTAGAGCAGCAG 2692

317 ............AspAlaArgLysMetGlnMetThrGlyAspLeuIleMe 329
2693 CTTCGCGAATCCGAGGCGCGCGCTGCGGAGCTGGCGAGCCAGCTG..... 2737

329 tLysAsnSerValThrValGlyGlyIleAlaGlyAlaSerGlyGlnTyrA 346
2738 .......................GAGGCCACTGCTGCTGCGAAGATGT 2762

346 laAlaThrGlnGluArgSer.................GluGlnGlnIle 356
2763 CAGCGGAGCAGGACCGCGAGAACACGAGGGCCACGCTAGAGCAGCAGCTT 2812

357 SerGlnValAsnAsnArgValAlaSerThrAlaSer.............. 368
2813 CGTGACTCCGAGGAGCGCGCTGCGGAGCTGGCGAGCCAGCTGGAGTCCAC 2862

369 .....................AspGluAlaArgGluSerSerArgLysS 378
2863 TACTGCTGCGAAGATGTCAGCGGAGCAGGACCGCGAGAGCACGAGG.... 2908

378 erThrSerLeuIleGlnGluMetLeuLysThrMetGluSerIleAsnGln 394
2909 ..GCCACGCTAGAGCAGCAGCTTCGTGACTCCGAGGAGCGCGCTGCGGAG 2956

395 SerLysAlaSerAlaLeuAlaAlaIleAlaGlyAsnIleArgAla*** 410
2957 CTGGCGAGCCAGCTGGAGTCCACTACTGCTGCGAAGATGTCAGCGGAG 3004
```

seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-282-845-1

```
seq_documentation_block:
; Sequence 1, Application US/08282845
; Patent No. 5719263
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven G.
;    TITLE OF INVENTION:  A 230Kd Antigen Present in Leishmania
;    TITLE OF INVENTION:  Species
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Immunex Corporation
;      STREET:  51 University Street
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98101
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Apple Macintosh
;      OPERATING SYSTEM:  Apple Macintosh Operating System 7.1
;      SOFTWARE:  Microsoft Word for Macintosh 5.1a
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/282,845
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/006,676
;      FILING DATE:  JANUARY 15, 1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Perkins, Patricia Anne
;      REGISTRATION NUMBER:  34,693
;      REFERENCE/DOCKET NUMBER:  5004-A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206)587-0430
;      TELEFAX:  (206)233-0644
```

```
OM of: US-09-068-804-6 to: Issued_Patents_NA:*   out_format : pfs

Date: Oct 5, 1999  7:10 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn2_1/USPTO_spool/US09068804/runat_05101999_145135_10326/app_query.fasta.1
 -DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=pam150
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=escore
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09068804
 -NCPU=6 -ICPU=3 -WAIT -THREADS=1

Search information block:
Query: US-09-068-804-6
Query length: 410
Database: Issued_Patents_NA:*
Database sequences: 192659
Database length: 52021692
Search time (sec): 101.650000
```

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	!	Documentation
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-006-676B-2	+	109.50	222.63	5.2e-05	3319	!	
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-282-845-1	+	109.50	222.63	5.2e-05	3319	!	
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-428-414A-4	+	109.50	222.63	5.2e-05	3319	!	
/cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:PCT-US94-00324-2	+	109.50	222.63	5.2e-05	3319		
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-038-682-1	+	102.50	200.25	0.0009	5116	!	
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-302-832-1	+	102.50	200.25	0.0009	5116	!	
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-530-198-1	+	102.50	200.25	0.0009	5116	!	
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-469-880-1	+	102.50	200.25	0.0009	5116	!	
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-728-470-1	+	102.50	200.25	0.0009	5116	!	
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-730-953-1	+	94.00	194.97	0.0018	1394	!	
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-038-682-5	+	102.50	193.25	0.0022	9171	!	
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-302-832-5	+	102.50	193.25	0.0022	9171	!	
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-530-198-5	+	102.50	193.25	0.0022	9171	!	
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-469-880-5	+	102.50	193.25	0.0022	9171	!	
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-728-470-5	+	102.50	193.25	0.0022	9171	!	
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-194-468-1	+	99.00	187.35	0.0048	7326	!	
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-038-682-7	+	95.00	183.96	0.0074	4287	!	
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-302-832-7	+	95.00	183.96	0.0074	4287	!	
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-530-198-7	+	95.00	183.96	0.0074	4287	!	
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-469-880-7	+	95.00	183.96	0.0074	4287	!	
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-728-470-7	+	95.00	183.96	0.0074	4287	!	
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-038-682-8	+	95.00	182.85	0.0085	4702	!	
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-302-832-8	+	95.00	182.85	0.0085	4702	!	
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-530-198-8	+	95.00	182.85	0.0085	4702	!	
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-469-880-8	+	95.00	182.85	0.0085	4702	!	
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-728-470-8	+	95.00	182.85	0.0085	4702	!	
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-387-156-9	+	87.50	177.10	0.0178	1635	!	
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-694-865-9	+	87.50	177.10	0.0178	1635	!	
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-843-530B-3	+	93.00	176.79	0.0186	5175	!	
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-682-517-13	+	90.50	174.79	0.0240	3666	!	
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-682-517-14	+	90.50	174.79	0.0240	3666	!	
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-843-530B-17	+	91.00	173.68	0.0276	4453	!	
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-737-526-3	-	83.00	173.50	0.0283	879	!	S
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-682-517-7	+	90.50	173.16	0.0295	4197	!	
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-682-517-8	+	90.50	173.16	0.0295	4197	!	
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-409-995-1	+	88.00	169.93	0.0447	3294	!	
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-694-865-15	+	84.00	165.50	0.0790	2102	!	
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-478-370-1	+	89.50	164.62	0.0884	6973	!	
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-620-605D-1	+	86.50	162.60	0.1145	4465	!	
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-568-459A-3	+	86.50	162.49	0.1162	4507	!	
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-09-005-232A-1	+	86.50	162.38	0.1178	4547	!	
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-461-775-8	+	80.00	161.25	0.1361	1320	!	
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-533-306A-1	+	81.00	158.97	0.1825	1960		
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-742-923A-1	+	81.00	158.97	0.1825	1960		
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-461-775-10	+	80.00	158.80	0.1865	1620		

```
seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-006-676B-2

seq_documentation_block:
; Sequence 2, Application US/08006676B
; Patent No. 5411865
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven
;    TITLE OF INVENTION:  Diagnosis of Leishmaniasis
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Jeffrey B. Oster
;      STREET:  8339 SE 57th Street
;      CITY:  Mercer Island
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98040-4906
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatable
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WORD for Windows
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/006,676B
;      FILING DATE:  15-JAN-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Oster, Jeffrey B.
;      REGISTRATION NUMBER:  32,585
;      REFERENCE/DOCKET NUMBER:  REED-4
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 232 7845
;      TELEFAX:  (206) 236 0205
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3319 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Leishmania chagasi
US-08-006-676B-2


alignment_scores:
            Quality:  109.50              Length:    516
              Ratio:  0.429                 Gaps:     22
 Percent Similarity:  49.419   Percent Identity: 19.574

alignment_block:
 US-09-068-804-6 x US-08-006-676B-2   ..

 Align seg 1/1  to: US-08-006-676B-2  from: 1  to: 3319

       2 LeuIleSerAsnValGlyIleAsnProAlaAlaTyrLeuAsnAsnHisSe 18
         :::|||:::::::|||:::   ::::::   :::|||
    1547 ATGATCGCGACTGTGAGCCCGAGCGCGCTGAACTAC.............. 1582

      18 rValGluAsnSerSerGlnThrAlaSerGlnSerValSerAlaLysAspI 35
           |||::::::     |||        :::      |||::|||||
    1583 ....GAGGAGACGCTGAGCACGCTGCGGTACGCGTCGCGCGCGCGCGACA 1628

      35 leLeuAsnSerIleGlyIleSer.............SerSerLysValSer 47
         ||:::|||        ::::::            :::   ::::::
    1629 TTGTGAATGTTGCGCAGGTGAACGAGGACCCGCGCGCACGGCGGATCCGC 1678

      48 AspLeu...............GlyLeuSerProThrLeuSerAlaProAl 59
         :::|||               :::::::   ::::::::::::::::   ::
    1679 GAGCTGGAGGAGCAGATGGAGGACATGCGGCAGGCGATGGCTGGCGGCGA 1728
```

```
     41 eSerSerSerLysValSerAspLeuGly......LeuSerProThrLeuS 56
        |||||||||      |||:::|||||||      ::::::   |||:::|
    422 CAGCAGCAGCGATGGCAGCAATCTTGGCATCATCATGACCGGAACTATAT 373

     56 erAlaProAlaProGlyValLeuThr............GlnThrProGly 68
        |||||   |||   :::|||   |||               |||||||||
    372 CCGCCGGAGCTGTATCCGTGGCCACCTCCACCGCCTCCGCCACCCCCGGA 323

     69 ThrIleThrSerPheLeuLysAlaSer..................... 77
        |||::::::      |||   ::::::
    322 ACTGTATCCGTGGTTCTCGGTGATAACAGCATAGGTGTTCACATGACCAC 273

     78 .............................................IleG 79
                                                       |||
    272 CGCCTCCTCCACCGCCGCTGTATCCGCCATCGTGGGAGTGAATCTCATAG 223

     79 lnAsnThrAspMetAsnGlnAspLeuAsnAla...LeuAlaAsnAsnVal 94
          :::|||:::|||:::      :::::|||   :::|||:::::::::
    222 CCTCCACCTCCATGGCCGCCAAAATCACCGCCTCCGTGGCCGCCAAAATC 173

     95 ThrThrLysAlaAsnGluValValGlnThrGlnLeuArgGluGlnGlnAl 111
        |||:::   :::::::::|||:::::|||   |||::::::   :::::
    172 ACCGCCTCCGTGGCCGCCGTGATCCCCACCGCTCTCTGGCAGGTGAATAC 123

    111 aGluValGlyLysPhePheAspIleSerGlyMetSerSer 124
        ::::|||:::::::::   :::   :::|||:::|||:::
    122 GGATGTGGACATGCTCTCGTCCGAAGGTGGGCTTTCCCCA 83
```

seq_name: gb_est3:H12259

seq_documentation_block:

```
LOCUS       H12259         528 bp    mRNA            EST       26-JUN-1995
DEFINITION  ym17d09.r1 Soares infant brain 1NIB Homo sapiens cDNA clone
            IMAGE:48274 3', mRNA sequence.
ACCESSION   H12259
NID         g877079
VERSION     H12259.1  GI:877079
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 528)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     On May 9, 1995 this sequence version replaced gi:803561.

            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 1700
            High quality sequence stops: 381
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Possible reversed clone: polyT not found
            Insert Length: 1700   Std Error: 0.00
            Seq primer: Promega -21m13
            High quality sequence stop: 381.
FEATURES             Location/Qualifiers
     source          1. .528
                     /organism="Homo sapiens"
                     /db_xref="GDB:420815"
                     /db_xref="taxon:9606"
                     /map="944E09; 12; 12q15-12q21.1"
                     /clone="IMAGE:48274"
                     /clone_lib="Soares infant brain 1NIB"
                     /sex="female"
                     /dev_stage="73 days post natal"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: whole brain; Vector: Lafmid BA; Site_1: Not
                     I; Site_2: Hind III; 1st strand cDNA was primed with a Not
                     I - oligo(dT) primer [5'
                     AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTT 3'];
                     double-stranded cDNA was ligated to Hind III adaptors
                     (Pharmacia), digested with Not I and directionally cloned
                     into the Not I and Hind III sites of the Lafmid BA vector.
                     Library went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      120 a    155 c    142 g    100 t     11 others
ORIGIN
```

alignment_scores:

```
           Quality:   78.50            Length:     90
             Ratio:  1.454              Gaps:      3
 Percent Similarity:  60.000   Percent Identity: 28.889
```

alignment_block:

US-09-068-804-6 x H12259 ..

Align seg 1/1 to: H12259 from: 1 to: 528

```
    167 MetArgGluGlyMetAsnAlaLeuSerGlySerIleSerGlnSerAlaLe 183
        |||   :::|||:::::::::|||||||||||:::::::|||:::::
     49 ATGGAGGCCGGTCTCTCTGAGCTGAGTGGCAGCGTGGCCCAGACAGTGAC 98

    183 uGlnLeuGlyIleThrGlyValGlyAlaLysLeuGluTyrLysGlyLeuG 200
         ||||||   :::|||                                |||
     99 TCAGTTACAGACGACC............................CTCG 118

    200 lnAsnGluArgGlyAlaLeuLysHisAsnAlaAlaLysIleAspLysLeu 216
          :::   :::      |||   :::::::::   |||||||:::   |||
    119 CCTCCGTCCAGGAGCTGCTGATTCAGCACAGCAGGAAGATCCAGGAGCTT 168

    217 ThrThrGluSerHisSerIleLysAsn.................ValLe 227
        :::   |||      :::   |||:::                   :::||
    169 GCCCACGAGCTGGCCGCTGCCAAGGCCACCACATCCACCAACTGGATCCT 218

    227 uAsnGlyGlnAsnSerValLysLeuGlyAlaGluGlyValAspSerLeuL 244
        |:::::|||||         |||   :::|||   :::::|||||||
    219 GGAGTCCCAGAATATCAACGAACTCAAGTCCGAA...ATTAACTCCTTGA 265

    244 ysSerLeuAsnMetLysLys 250
        ||:::|||   :::::::::
    266 AAGGGCTTCTTTTAAATCGG 285
```

```
          Washington University School of MedicineP
          4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
          Tel: 314 286 1800
          Fax: 314 286 1810
          Email: mouseest@watson.wustl.edu
          This clone is available royalty-free through LLNL ; contact the
          IMAGE Consortium (info@image.llnl.gov) for further information.
          MGI:653397
          Seq primer: -28m13 rev1 ET from Amersham
          High quality sequence stop: 288.
FEATURES             Location/Qualifiers
     source          1. .374
                     /organism="Mus musculus"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1227805"
                     /clone_lib="Stratagene mouse skin (#937313)"
                     /sex="females"
                     /tissue_type="whole skin"
                     /dev_stage="11 weeks old"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Organ: skin; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. Whole skin from 11 week old C57BL/6 female mice.
                     Average insert size: 1.0 kb; Uni-ZAP XR Vector; ~5'
                     adaptor sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor
                     sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT       90 a     91 c    130 g     63 t
ORIGIN


alignment_scores:
            Quality:   77.00             Length:    169
              Ratio:  0.906               Gaps:      6
 Percent Similarity:  50.296   Percent Identity: 21.893

alignment_block:
 US-09-068-804-6 x AA760238/rev   ..

 Align seg 1/1  to reverse of: AA760238  from: 1  to: 374

      58 ProAlaProGlyValLeuThrGlnThr.........ProGlyThrIleTh 71
         |||:::      :::||||||||||||         |||   ::::::::
     373 CCTCCACGGGAGCTGCTGACACAGACATTCACAGAGCCCACACCCCTCGC 324

      71 rSerPheLeuLysAlaSerIleGlnAsnThrAspMetAsnGlnAspLeuA 88
         :      |||     ::::::|||::::::   ::::::::|||||||:
     323 AGAGCCTCTGCTCCTCGCCCTCCAGCAGGCGCCTGTAGGTGATGATCTCG 274

      88 snAlaLeuAlaAsnAsnValThrThrLysAlaAsnGluValValGlnThr 104
         ::                                        :::|||:::
     273 AC.....................................ATCCAGCCC 263

     105 GlnLeuArgGluGlnGlnAlaGluValGlyLysPhePheAspIleSerGl 121
         ||||||      ::::::
     262 CAACTTGGAGTTCATCAC.............................. 245

     121 yMetSerSerSerAlaValAlaLeuLeuAlaAlaAlaAsnThrLeuMetL 138
                      :::::::::|||:::   |||:::   ::::::|
     244 .............CTCCTGGTATTCCTTGAGCAGGCAGGCCATATCCTGC 208

     138 euThrLeuAsnGlnAlaAspSerLysLeuSerGlyLysLeuSerLeuVal 154
         ||:::|||                 |||   |||::::||   |||
     207 TTAGCCTT.................CTGCAGGGCACCCTCCAGCTCAGC 176

     155 SerPheAspAlaAlaLysThrThrAlaSerSerMetMetArgGluGlyMe 171
            ::::::|||:::                      :::   ||||||
     175 CAGCTTGCAGCGGGC...................ATCAGTGAGGGCAG 147

     171 tAsnAlaLeuSerGlySerIleSerGlnSerAlaLeuGlnLeuGlyIleT 188
            :::|||        :::::::::|||:::||||||||||||:::
     146 CCTCTCCCTGTTGCTCAGATTGGGTCACAGCAGCCTCCAGCTTGGTGTTC 97

     188 hrGlyValGlyAlaLysLeuGluTyrLysGlyLeuGlnAsnGluArgGly 204
         :::::|||                    |||:::
      96 TGGCACTTGGCATT.................CTCGATCTCAGCAGTCAG 65

     205 AlaLeuLysHisAsnAlaAlaLysIleAspLysLeuThrThrGluSerHi 221
         ::::||   |||::::::      ||||||   |||:::   ::::::::
      64 CCTCTGGATCATTCTGTTCAGCTCATTGATCTCCTCTCTGGTGCGGCGCA 15

     221 sSerIle 223
         :::::::
      14 GGGTCTC 8

seq_name: gb_est14:AA391538

seq_documentation_block:
LOCUS       AA391538      474 bp    mRNA            EST       28-NOV-1998
DEFINITION  LD10548.5prime LD Drosophila melanogaster embryo BlueScript
            Drosophila melanogaster cDNA clone LD10548 5prime, mRNA sequence.
ACCESSION   AA391538
NID         g2044522
VERSION     AA391538.1  GI:2044522
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 474)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1397762.

            Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 105  row: D  column: 12
            High quality sequence stop: 463.
FEATURES             Location/Qualifiers
     source          1. .474
                     /organism="Drosophila melanogaster"
                     /db_xref="BDGP_EST:BDcln009803"
                     /db_xref="taxon:7227"
                     /clone="LD10548"
                     /clone_lib="LD Drosophila melanogaster embryo BlueScript"
                     /sex="male and female"
                     /dev_stage="0 to 24 hours mixed stage embryonic"
                     /lab_host="SOLR"
                     /note="Organ: embryo; Vector: BlueScript SK; Site_1:
                     EcoRI; Site_2: XhoI; Constructed using Stratagene ZAP-cDNA
                     Synthesis kit. Oligo dT-primed and directionally cloned at
                     EcoRI and XhoI in BlueScript SK(+/-)"
BASE COUNT       96 a    120 c    166 g     92 t
ORIGIN


alignment_scores:
            Quality:   78.00             Length:    130
              Ratio:  0.987               Gaps:      4
 Percent Similarity:  60.769   Percent Identity: 25.385

alignment_block:
 US-09-068-804-6 x AA391538/rev   ..

 Align seg 1/1  to reverse of: AA391538  from: 1  to: 474

      25 ThrAlaSerGlnSerValSerAlaLysAspIleLeuAsnSerIleGlyIl 41
         :::|||:::   |||:::||||||:::::::::::::::::   :::
     472 TCCGCCGCTGCCTCCATGTCCGCCATGTCCGTGATCGGAGGACGAGCCAG 423
```

```
     225 AsnValLeuAsn...GlyGlnAsnSerValLysLeuGlyAlaGluGlyVa 240
         ||||||||||||   :::::::::||||||:::::::::::::      ::
      51 AACGTCTTGAATTGGAACGAATCGAGCGTGAACGTCGCGAGCTGGAAGCT 100

     240 lAspSerLeuLysSerLeu......AsnMetLysLysThrGlyThrAspA 255
         :::::|||:::::::::|||      |||:::  |||      |||:::
     101 CGAGAGCGTCAGGAACTTGAGCTTCAACGTCGTGAAGCTGAAGACCGGGA 150

     255 laThrLysAsnLeuAsnAspAlaThr......LeuLysSerAsnAlaGly 269
           :::  |||:::|||:::::::      ::::::      :::
     151 AACGCCAACGTCTCGAAGATGAGGCTCGCGAGATGCGCCGTCGTGAAGAA 200

     270 ThrSerAlaThrGluSerLeuGlyIleLysAspSerAsnLysGlnIleSe 286
                 :::   |||   :::   |||   ::::::|||
     201 GAGCGCCGAGAAGCAGAATTG...GTTGCAGATGTCCATAGACAA..... 242

     286 rProGluHisGlnAlaIleLeuSerLysArgLeuGluSerValGluSerA 303
         :::|||:::::::::   |||   ||||||   |||      |||::::
     243 .GCTGAAGAACGTGAACGACTCCGCAAGCGTCAAGAGCGTGAAGAGGCTG 291

     303 sp............IleArgLeuGluGlnAsnThrMetAspMetThrArg 315
         ::            ||||||||||||||||:::::::::||||||   |||
     292 AACGTTTGGAAAGAATCCGTCTTGAGCAGCAGAAGATTGACATGGAAAGA 341

     316 IleAspAlaArgLys 320
         |||||||||   :::
     342 ATTGACGCTGAACGA 356

seq_name: gb_est28:AI516432

seq_documentation_block:
LOCUS       AI516432      547 bp    mRNA            EST       16-MAR-1999
DEFINITION  LD42296.5prime LD Drosophila melanogaster embryo pOT2 Drosophila
            melanogaster cDNA clone LD42296 5prime, mRNA sequence.
ACCESSION   AI516432
NID         g4419532
VERSION     AI516432.1  GI:4419532
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 547)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     On Mar 10, 1998 this sequence version replaced gi:2949298.

            Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            hit genomic sequence AC005709; hit genomic sequence AC006243
            Plate: 422  row: H  column: 12
            High quality sequence stop: 455.
FEATURES             Location/Qualifiers
     source          1. .547
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="LD42296"
                     /clone_lib="LD Drosophila melanogaster embryo pOT2"
                     /sex="male and female"
                     /dev_stage="0 to 24 hours mixed stage embryonic"
                     /lab_host="XL1 Blue"
                     /note="Organ: embryo; Vector: pOT2; Site_1: EcoRI; Site_2:
                     XhoI; Sized fractionated cDNAs were directly ligated into
                     pOT2. "
BASE COUNT      112 a    138 c    189 g    108 t

ORIGIN


alignment_scores:
            Quality:   79.00            Length:    133
              Ratio:  0.975              Gaps:      4
 Percent Similarity:  60.902   Percent Identity: 25.564

alignment_block:
 US-09-068-804-6 x AI516432/rev   ..

 Align seg 1/1  to reverse of: AI516432  from: 1  to: 547

      22 SerSerGlnThrAlaSerGlnSerValSerAlaLysAspIleLeuAsnSe 38
         |||:::   :::|||:::   |||:::||||||:::::::::::::::::
     540 TCCGCCGTGTCCGCCGCTGCCTCCATGTCCGCCATGTCCGTGATCGGAGG 491

      38 rIleGlyIleSerSerSerLysValSerAspLeuGly......LeuSerP 53
         :   :::   |||||||||      |||:::|||||      ::::::
     490 ACGAGCCAGCAGCAGCAGCGATGGCAGCAATCTTGGCATCATCATGACCG 441

      53 roThrLeuSerAlaProAlaProGlyValLeuThr............Gln 65
           |||:::||||||   |||   :::|||   |||
     440 GAACTATATCCGCCGGAGCTGTATCCGTGGCCACCTCCACCGCCTCCGCC 391

      66 ThrProGlyThrIleThrSerPheLeuLysAlaSer.............. 77
         ||||||||||||::::::      |||   ::::::
     390 ACCCCCGGAACTGTATCCGTGGTTCTCGGTGATAACAGCATAGGTGTTCA 341

      77 ..................................................... 77

     340 CATGACCACCGCCTCCTCCACCGCCGCTGTATCCGCCATCGTGGGAGTGA 291

      78 .....IleGlnAsnThrAspMetAsnGlnAspLeuAsnAla...LeuAla 91
           |||   :::|||:::|||:::      ::::::|||   :::|||
     290 ATCTCATAGCCTCCACCTCCATGGCCGCCAAAATCACCGCCTCCGTGGCC 241

      92 AsnAsnValThrThrLysAlaAsnGluValValGlnThrGlnLeuArgGl 108
         :::::::::|||:::   ::::::::::|||:::::::|||   |||::::
     240 GCCAAAATCACCGCCTCCGTGGCCGCCGTGATCCCCACCGCTCTCTGGCA 191

     108 uGlnGlnAlaGluValGlyLysPhePheAspIleSerGlyMetSerSer 124
         :   ::::::::::|||::::::::::   :::   :::|||:::|||:::
     190 GGTGAATACGGATGTGGACATGCTCTCGTCCGAAGGTGGGCTTTCCCCA 142

seq_name: gb_est19:AA760238

seq_documentation_block:
LOCUS       AA760238      374 bp    mRNA            EST       23-JAN-1998
DEFINITION  vv70h07.r1 Stratagene mouse skin (#937313) Mus musculus cDNA clone
            IMAGE:1227805 5' similar to gb:M21389 KERATIN, TYPE II CYTOSKELETAL
            5 (HUMAN); gb:M92088 Mus musculus type II hair keratin mRNA, 3' end
            (MOUSE);, mRNA sequence.
ACCESSION   AA760238
NID         g2808032
VERSION     AA760238.1  GI:2808032
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 374)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1395353.

            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
```

```
    220 rHisSerIle 223
        ::::::::::
     74 GCAGGGTCTC 65

seq_name: gb_est25:AI295460

seq_documentation_block:
LOCUS       AI295460      536 bp    mRNA            EST       01-DEC-1998
DEFINITION  LP09102.5prime LP Drosophila melanogaster larval-early pupal pOT2
            Drosophila melanogaster cDNA clone LP09102 5prime, mRNA sequence.
ACCESSION   AI295460
NID         g3944867
VERSION     AI295460.1  GI:3944867
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 536)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 17, 1998 this sequence version replaced gi:2043255.

            Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 91  row: A  column: 2
            High quality sequence stop: 420.
FEATURES             Location/Qualifiers
     source          1. .536
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="LP09102"
                     /clone_lib="LP Drosophila melanogaster larval-early pupal
                     pOT2"
                     /sex="male and female"
                     /dev_stage="larvae-pupae"
                     /lab_host="DH5-alpha"
                     /note="Organ: whole body; Vector: pOT2; Site_1: EcoRI;
                     Site_2: Xho1; Sized fractionated cDNAs were directly
                     ligated into pOT2. Plasmid cDNA library.
                                 "
BASE COUNT      111 a    138 c    185 g    102 t
ORIGIN


alignment_scores:
            Quality:   79.00             Length:    133
              Ratio:  0.975               Gaps:      4
 Percent Similarity:  60.902   Percent Identity: 25.564

alignment_block:
 US-09-068-804-6 x AI295460/rev   ..

 Align seg 1/1  to reverse of: AI295460  from: 1  to: 536

      22 SerSerGlnThrAlaSerGlnSerValSerAlaLysAspIleLeuAsnSe 38
         |||:::   :::|||:::   |||:::||||||:::::::::::::::::
     497 TCCGCCGTGTCCGCCGCTGCCTCCATGTCCGCCATGTCCGTGATCGGAGG 448

      38 rIleGlyIleSerSerSerLysValSerAspLeuGly......LeuSerP 53
         :   :::   |||||||||      |||:::||||||      ::::::
     447 ACGAGCCAGCAGCAGCAGCGATGGCAGCAATCTTGGCATCATCATGACCG 398

      53 roThrLeuSerAlaProAlaProGlyValLeuThr.....Gln 65
       .  . :|||:::||||||   |||   :::|||   |||

     397 GAACTATATCCGCCGGAGCTGTATCCGTGGCCACCTCCACCGCCTCCGCC 348

      66 ThrProGlyThrIleThrSerPheLeuLysAlaSer..............  77
         ||||||||||||:::::::      |||   ::::::
     347 ACCCCCGGAACTGTATCCGTGGTTCTCGGTGATAACAGCATAGGTGTTCA 298

      77 ..............................................  77

     297 CATGACCACCGCCTCCTCCACCGCCGCTGTATCCGCCATCGTGGGAGTGA 248

      78 .....IleGlnAsnThrAspMetAsnGlnAspLeuAsnAla...LeuAla 91
            |||   :::|||:::|||:::      ::::::|||   :::|||
     247 ATCTCATAGCCTCCACCTCCATGGCCGCCAAAATCACCGCCTCCGTGGCC 198

      92 AsnAsnValThrThrLysAlaAsnGluValValGlnThrGlnLeuArgGl 108
         :::::::::|||:::   ::::::::::|||::::::|||   |||:::::
     197 GCCAAAATCACCGCCTCCGTGGCCGCCGTGATCCCCACCGCTCTCTGGCA 148

     108 uGlnGlnAlaGluValGlyLysPhePheAspIleSerGlyMetSerSer 124
         :   :::::::::|||::::::::::   :::   :::|||:::|||:::
     147 GGTGAATACGGATGTGGACATGCTCTCGTCCGAAGGTGGGCTTTCCCCA 99

seq_name: gb_est17:C63675

seq_documentation_block:
LOCUS       C63675        360 bp    mRNA            EST       22-SEP-1997
DEFINITION  C63675 Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
            clone yk348a11 5', mRNA sequence.
ACCESSION   C63675
NID         g2422380
VERSION     C63675.1  GI:2422380
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 360)
  AUTHORS   Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,
            Sano,M., Miyata,A. and Nishigaki,A.
  TITLE     Expression map of the C.elegans genome
  JOURNAL   Unpublished (1996)
COMMENT     On Apr 14, 1993 this sequence version replaced gi:693089.

            Contact: Yuji Kohara
            Gene Library Lab
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411, Japan
            Tel: 0559-75-0771
            Fax: 0559-75-6240
            Email: ykohara@ddbj.nig.ac.j.
FEATURES             Location/Qualifiers
     source          1. .360
                     /organism="Caenorhabditis elegans"
                     /strain="CB1489 him-8(e1489)"
                     /db_xref="taxon:6239"
                     /map="3; 808C09; 2; 2q22.1-2q23.3"
                     /clone="yk348a11"
                     /clone_lib="Yuji Kohara unpublished cDNA"
                     /note="dev_stage=varied, sex=Hermaphrodite male,
                     tissue_type=whole animal"
BASE COUNT      110 a     73 c    119 g     58 t
ORIGIN


alignment_scores:
            Quality:  77.00             Length:    105
              Ratio:  1.132               Gaps:      6
 Percent Similarity:  64.762   Percent Identity: 31.429

alignment_block:
 US-09-068-804-6 x C63675   ..

 Align seg 1/1  to: C63675  from: 1  to: 360
```

```
alignment_block:
 US-09-068-804-6 x AI294100/rev   ..

 Align seg 1/1  to reverse of: AI294100  from: 1  to: 525

      22 SerSerGlnThrAlaSerGlnSerValSerAlaLysAspIleLeuAsnSe 38
         |||:::   :::|||:::   |||:::||||||:::::::::::::::::
     486 TCCGCCGTGTCCGCCGCTGCCTCCATGTCCGCCATGTCCGTGATCGGAGG 437

      38 rIleGlyIleSerSerSerLysValSerAspLeuGly......LeuSerP 53
         :   :::   ||||||||:      |||:::|||||:      ::::::
     436 ACGAGCCAGCAGCAGCAGCGATGGCAGCAATCTTGGCATCATCATGACCG 387

      53 roThrLeuSerAlaProAlaProGlyValLeuThr............Gln 65
           |||:::||||||   |||   :::|||   |||
     386 GAACTATATCCGCCGGAGCTGTATCCGTGGCCACCTCCACCGCCTCCGCC 337

      66 ThrProGlyThrIleThrSerPheLeuLysAlaSer.............. 77
         ||||||||||||:::::      |||   ::::::
     336 ACCCCCGGAACTGTATCCGTGGTTCTCGGTGATAACAGCATAGGTGTTCA 287

      77 ..................................................... 77

     286 CATGACCACCGCCTCCTCCACCGCCGCTGTATCCGCCATCGTGGGAGTGA 237

      78 .....IleGlnAsnThrAspMetAsnGlnAspLeuAsnAla...LeuAla 91
            |||   :::|||:::|||:::      ::::::|||   :::|||
     236 ATCTCATAGCCTCCACCTCCATGGCCGCCAAAATCACCGCCTCCGTGGCC 187

      92 AsnAsnValThrThrLysAlaAsnGluValValGlnThrGlnLeuArgGl 108
         ::::::::::|||:::   ::::::::::|||::::::|||   |||:::::
     186 GCCAAAATCACCGCCTCCGTGGCCGCCGTGATCCCCACCGCTCTCTGGCA 137

     108 uGlnGlnAlaGluValGlyLysPhePheAspIleSerGlyMetSerSer 124
         :   ::::::::|||::::::::::   :::   :::|||:::|||:::
     136 GGTGAATACGGATGTGGACATGCTCTCGTCCGAAGGTGGGCTTTCCCCA 88

seq_name: gb_est16:AA560346

seq_documentation_block:
LOCUS       AA560346      433 bp    mRNA            EST       18-AUG-1997
DEFINITION  vl39a10.r1 Stratagene mouse skin (#937313) Mus musculus cDNA clone
            IMAGE:974586 5' similar to gb:M21389 KERATIN, TYPE II CYTOSKELETAL
            5 (HUMAN); gb:M92088 Mus musculus type II hair keratin mRNA, 3' end
            (MOUSE);, mRNA sequence.
ACCESSION   AA560346
NID         g2331811
VERSION     AA560346.1  GI:2331811
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 433)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On May 18, 1995 this sequence version replaced gi:811348.

            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:555314

            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 355.
FEATURES             Location/Qualifiers
     source          1. .433
                     /organism="Mus musculus"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /map="22"
                     /clone="IMAGE:974586"
                     /clone_lib="Stratagene mouse skin (#937313)"
                     /sex="females"
                     /tissue_type="whole skin"
                     /dev_stage="11 weeks old"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Organ: skin; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. Whole skin from 11 week old C57BL/6 female mice.
                     Average insert size: 1.0 kb; Uni-ZAP XR Vector; ~5'
                     adaptor sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor
                     sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT      104 a    104 c    150 g     74 t      1 others
ORIGIN


alignment_scores:
            Quality:   78.00              Length:    170
              Ratio:  0.918                Gaps:      6
 Percent Similarity:  50.000   Percent Identity: 22.353

alignment_block:
 US-09-068-804-6 x AA560346/rev   ..

 Align seg 1/1  to reverse of: AA560346  from: 1  to: 433

      57 AlaProAlaProGlyValLeuThrGlnThr.........ProGlyThrIl 70
         ::::|||:::      :::|||||||||||||        |||   :::
     433 ACTCCTCCACGGGAGCTGCTGACACAGACATTCACAGAGCCCACACCCTC 384

      70 eThrSerPheLeuLysAlaSerIleGlnAsnThrAspMetAsnGlnAspL 87
            |||   |||      ::::::|||::::::   :::::::::||||
     383 GCAGAGCNCTCTGCTCCTCGCCCTCCAGCAGGCGCCTGTAGGTGATGATC 334

      87 euAsnAlaLeuAlaAsnAsnValThrThrLysAlaAsnGluValValGln 103
         ||:::                                       :::|||
     333 TCGAC.....................................ATCCAG 323

     104 ThrGlnLeuArgGluGlnGlnAlaGluValGlyLysPhePheAspIleSe 120
         :::||||||      ::::::
     322 CCCCAACTTGGAGTTCATCAC..................... 302

     120 rGlyMetSerSerSerAlaValAlaLeuLeuAlaAlaAlaAsnThrLeuM 137
                        ::::::::::|||:::   |||:::   ::::
     301 ................CTCCTGGTATTCCTTGAGCAGGCAGGCCATATCC 268

     137 etLeuThrLeuAsnGlnAlaAspSerLysLeuSerGlyLysLeuSerLeu 153
         ::|||:::|||                |||   |||:::|||   |||
     267 TGCTTAGCCTT.................CTGCAGGGCACCCTCCAGCTC 236

     154 ValSerPheAspAlaAlaLysThrThrAlaSerSerMetMetArgGluGl 170
               ::::::|||:::                        :::   |||||
     235 AGCCAGCTTGCAGCGGGC.....................ATCAGTGAGGG 207

     170 yMetAsnAlaLeuSerGlySerIleSerGlnSerAlaLeuGlnLeuGlyI 187
         |      :::|||         ::::::::::|||:::||||||||||||:
     206 CAGCCTCTCCCTGTTGCTCAGATTGGGTCACAGCAGCCTCCAGCTTGGTG 157

     187 leThrGlyValGlyAlaLysLeuGluTyrLysGlyLeuGlnAsnGluArg 203
         ::   ::::::|||                    |||:::
     156 TTCTGGCACTTGGCATT.................CTCGATCTCAGCAGT 125

     204 GlyAlaLeuLysHisAsnAlaAlaLysIleAspLysLeuThrThrGluSe 220
           :::|||   |||::::::      ||||||   |||:::   :::::
     124 CAGCCTCTGGATCATTCTGTTCAGCTCATTGATCTCCTCTCTGGTGCGGC 75
```

```
            discoideum cDNA clone SLB724, mRNA sequence.
ACCESSION   AU034008
NID         g3799432
VERSION     AU034008.1  GI:3799432
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 688)
  AUTHORS   Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
            Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
            Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
  TITLE     The Dictyostelium developmental cDNA project: generation and
            analysis of expressed sequence tags from the first-finger stage of
            development
  JOURNAL   DNA Res. 5 (6), 335-340 (1998)
  MEDLINE   99156227
COMMENT     On Jan 14, 1998 this sequence version replaced gi:1797180.

            Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan
            Email: d402hu@sakura.cc.tsukuba.ac.jp
            PROJECT = 'Dictyostelium discoideum cDNA project in Japan'
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .688
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="SLB724"
                     /clone_lib="Dictyostelium discoideum SL (H.Urushihara)"
                     /dev_stage="slug"
BASE COUNT      252 a     98 c    104 g    234 t
ORIGIN


alignment_scores:
            Quality:   80.50            Length:    170
              Ratio:  0.856              Gaps:      6
 Percent Similarity:  55.294   Percent Identity: 21.765

alignment_block:
 US-09-068-804-6 x AU034008   ..

 Align seg 1/1  to: AU034008  from: 1  to: 688

     223 IleLysAsnValLeuAsnGlyGlnAsnSerValLysLeuGlyAlaGluGl 239
         |||:::::::   |||:::|||:::::::::|||||:::      |||||
      56 ATCAATGCCCAATTGGATGGCCATGGTGGTGTTAAATTCACTAGAGAAGG 105

     239 y..................ValAspSerLeuLysSerLeuAsnMetLysL 250
         |                  :::|||   |||   |||::::::::::|||:
     106 TGATTTCCAAACTTCATATTTAGATAAACTT...AGCTTTGATATTAAAA 152

     250 ysThr.............................GlyThrAspAlaThr 256
         :::::                           |||      ||||||
     153 CCTCTGCTGATAGTAGTTCATTTGCAGTTTATTTTGGTGATGTTGCAACT 202

     257 LysAsnLeuAsnAspAlaThrLeuLysSerAsnAlaGlyThrSerAlaTh 273
         ||||||::::::      :::            |||:::|||   :::||
     203 AAAAACATTGACCCATTAGTA............GCCTCAACTTGGACCAC 240

     273 rGluSerLeuGlyIleLysAspSerAsnLysGlnIleSerProGluHisG 290
         |     |||::::::|||:::   ::::::::::::
     241 TGTTGAATTGAGTGTTAAATCTTTAACAAATAATACAGTTGAAGGTTCTA 290

     290 lnAlaIleLeuSerLysArgLeuGluSerValGlu.............. 301
           :::::::::::::::::   ::::::::::::
     291 TAACATTTTTCAATAATCAAGCTGATCCAATTAATATATGGTTAGATAAT 340

     302 .g.................SerAspIleArgLeuGluGlnAsnThrMetAs 312
                           :::|||      |||:::|||   ||||||||
     341 ATTAAATGGACATACACAGATGATGCTCCTCTTGATCAACCAACAATGGA 390

     312 pMetThrArgIleAspAlaArgLysMetGlnMetThrGlyAspLeuIleM 329
         |::::::   ::::::::::::            :::|||      :::
     391 TTTAATTGAAGTAAATAATAAACCATCAACAACATCAGGTACTGGAACCA 440

     329 etLysAsn.....................SerValThrValGlyGlyIle 338
           :::                       :::|||:::   ||||||||
     441 CTTCAAGTAAACCTTCATCTAGTAGTGGTGGTGTTAGTGGTGGTGGTATT 490

     339 AlaGlyAlaSerGlyGlnTyrAlaAlaThrGlnGluArgSerGluGlnGl 355
         ::::::::::|||:::   |||:::   :::   |||      |||:::||
     491 GGTAGTGAAAGTTCAATCTATGGATTAAGTTCTGAAAACCAAGAAAATCA 540

     355 nIleSerGln 358
         |   ::::::
     541 AAGTGGACAT 550

seq_name: gb_est25:AI294100

seq_documentation_block:
LOCUS       AI294100      525 bp    mRNA            EST       01-DEC-1998
DEFINITION  LP07402.5prime LP Drosophila melanogaster larval-early pupal pOT2
            Drosophila melanogaster cDNA clone LP07402 5prime, mRNA sequence.
ACCESSION   AI294100
NID         g3943507
VERSION     AI294100.1  GI:3943507
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 525)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     On Oct 17, 1997 this sequence version replaced gi:2519259.

            Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 74  row: A  column: 2
            High quality sequence stop: 485.
FEATURES             Location/Qualifiers
     source          1. .525
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /map="18,11"
                     /clone="LP07402"
                     /clone_lib="LP Drosophila melanogaster larval-early pupal
                     pOT2"
                     /sex="male and female"
                     /dev_stage="larvae-pupae"
                     /lab_host="DH5-alpha"
                     /note="Organ: whole body; Vector: pOT2; Site_1: EcoRI;
                     Site_2: Xho1; Sized fractionated cDNAs were directly
                     ligated into pOT2. Plasmid cDNA library.
                                "
BASE COUNT      108 a    135 c    179 g    103 t
ORIGIN


alignment_scores:
            Quality:   79.00            Length:    133
              Ratio:  0.975              Gaps:      4
 Percent Similarity:  60.902   Percent Identity: 25.564
```

```
ORIGIN


alignment_scores:
           Quality:   79.00              Length:    128
             Ratio:  1.026               Gaps:        4
 Percent Similarity:  60.156   Percent Identity: 18.750

alignment_block:
 US-09-068-804-6 x AI372258   ..

 Align seg 1/1  to: AI372258  from: 1  to: 416

     298 GluSerValGluSerAspIleArgLeuGluGlnAsnThrMetAspMetTh 314
         ::::::::||   ::::::::::|||:::      ::::::::::::::::
       2 GACGCCGTCTCCGCTCAAGTGAGGTTCGGAGCTCAGAGCCTTGCGATCTC 51

     314 rArgIleAspAlaArgLysMet.......................... 321
         :|||:::   ||||||||||||
      52 TCGGCTGCGCGCGCGCAAGATGGCTGACCCAGAATTGGAAGCTATCAGGA 101

     322 ..GlnMetThrGlyAspLeuIleMetLysAsnSerValThrValGlyGly 337
           :::          :::|||:::   |||:::::::
     102 AGAGAAGGATGGAAGAACTAATGGCAAAACACGGT.............. 136

     338 IleAlaGlyAlaSerGlyGlnTyrAlaAlaThrGlnGluArgSerGluGl 354
         :::|||:::::::::|||   :::|||:::|||:::::::::::::::
     137 ...GGCGGTGGGGCAAATCAACAAAATGCAAGCCAACAAAAAGCTCAGGA 183

     354 nGlnIleSerGlnValAsnAsn..............ArgValAlaSerT 366
         ::::      |||   ::::::                :::|||   :
     184 AGACGCGAAACAGGAAGCTGAGGAACGCCGACAGATGATGCTTGCCCAGA 233

     366 hrAlaSerAspGluAlaArgGluSerSerArgLysSerThrSerLeuIle 382
         ::   |||:::||||||||||||          :::   :::   :::
     234 TATTATCTTCTGAAGCCAGAGAAAGGATCTCCCGCATAGCTTTGGTCAAA 283

     383 GlnGluMetLeuLysThrMetGluSerIleAsnGlnSerLysAlaSerAl 399
         ::::::::::   :::   :::|||::::::          :::|||:::
     284 CCTGACAAAGCAAGAGGGGTGGAGGATGTT...CTGCTGAGAGCTGCTCA 330

     399 aLeuAlaAlaIleAlaGlyAsnIleArgAla*** 410
           :::::::|||:::   ::::::   :::|||
     331 AACTGGTGGAATATCTGAGAAGGTGTCCGAAGAA 364

seq_name: gb_est25:AU039285

seq_documentation_block:
LOCUS       AU039285      643 bp    mRNA            EST       29-MAR-1999
DEFINITION  AU039285 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium
            discoideum cDNA clone SLH350, mRNA sequence.
ACCESSION   AU039285
NID         g4008266
VERSION     AU039285.1  GI:4008266
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Dictyostelida; Dictyostelium.
REFERENCE   1  (bases 1 to 643)
  AUTHORS   Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
            Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
            Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
  TITLE     The Dictyostelium developmental cDNA project: generation and
            analysis of expressed sequence tags from the first-finger stage of
            development
  JOURNAL   DNA Res. 5 (6), 335-340 (1998)
  MEDLINE   99156227
COMMENT     On May 8, 1995 this sequence version replaced gi:801091.

            Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan
            Email: d402hu@sakura.cc.tsukuba.ac.jp
            PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES             Location/Qualifiers
     source          1. .643
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="SLH350"
                     /clone_lib="Dictyostelium discoideum SL (H.Urushihara)"
                     /dev_stage="slug"
BASE COUNT      235 a     93 c     94 g    221 t
ORIGIN


alignment_scores:
           Quality:   80.50              Length:    170
             Ratio:  0.856               Gaps:        6
 Percent Similarity:  55.294   Percent Identity: 21.765

alignment_block:
 US-09-068-804-6 x AU039285   ..

 Align seg 1/1  to: AU039285  from: 1  to: 643

     223 IleLysAsnValLeuAsnGlyGlnAsnSerValLysLeuGlyAlaGluGl 239
         |||::::::   |||:::|||:::::::::|||||:::      |||||
       6 ATCAATGCCCAATTGGATGGCCATGGTGGTGTTAAATTCACTAGAGAAGG 55

     239 y.................ValAspSerLeuLysSerLeuAsnMetLysL 250
         |                  :::|||   |||   |||::::::::::|||:
      56 TGATTTCCAAACTTCATATTTAGATAAACTT...AGCTTTGATATTAAAA 102

     250 ysThr............................GlyThrAspAlaThr 256
         :::::                           |||      ||||||
     103 CCTCTGCTGATAGTAGTTCATTTGCAGTTTATTTTGGTGATGTTGCAACT 152

     257 LysAsnLeuAsnAspAlaThrLeuLysSerAsnAlaGlyThrSerAlaTh 273
         ||||||::::::      :::           |||:::|||   :::||
     153 AAAAACATTGACCCATTAGTA.............GCCTCAACTTGGACCAC 190

     273 rGluSerLeuGlyIleLysAspSerAsnLysGlnIleSerProGluHisG 290
         |      |||:::::::|||:::   ::::::::::::
     191 TGTTGAATTGAGTGTTAAATCTTTAACAAATAATACAGTTGAAGGTTCTA 240

     290 lnAlaIleLeuSerLysArgLeuGluSerValGlu.............. 301
           ::::::::::::::::::   ::::::::::::
     241 TAACATTTTTCAATAATCAAGCTGATCCAATTAATATATGGTTAGATAAT 290

     302 ..................SerAspIleArgLeuGluGlnAsnThrMetAs 312
                            :::|||      |||:::|||   ||||||||
     291 ATTAAATGGACATACACAGATGATGCTCCTCTTGATCAACCAACAATGGA 340

     312 pMetThrArgIleAspAlaArgLysMetGlnMetThrGlyAspLeuIleM 329
         |::::::   ::::::::::::            :::|||      :::
     341 TTTAATTGAAGTAAATAATAAACCATCAACAACATCAGGTACTGGAACCA 390

     329 etLysAsn.....................SerValThrValGlyGlyIle 338
           :::                       :::|||:::   ||||||||
     391 CTTCAAGTAAACCTTCATCTAGTAGTGGTGGTGTTAGTGGTGGTGGTATT 440

     339 AlaGlyAlaSerGlyGlnTyrAlaAlaThrGlnGluArgSerGluGlnGl 355
         :::::::::|||:::   |||:::   :::   |||      |||:::||
     441 GGTAGTGAAAGTTCAATCTATGGATTAAGTTCTGAAAACCAAGAAAATCA 490

     355 nIleSerGln 358
         |   ::::::
     491 AAGTGGACAT 500

seq_name: gb_est24:AU034008

seq_documentation_block:
LOCUS       AU034008      688 bp    mRNA            EST       28-APR-1999
DEFINITION  AU034008 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium
```

```
          300 Herston Road, Qeensland 4029, Australia
          Tel: 61 7 3362 0413
          Fax: 61 7 3362 0104
          Email: paulB@qimr.edu.au
          PCR PRimers
          FORWARD: M13 Forward
          BACKWARD: M13 Reverse
          Insert Length: 1000   Std Error: 0.00
          Seq primer: T3 Reverse
          High quality sequence stop: 399.
FEATURES             Location/Qualifiers
     source          1. .399
                     /organism="Schistosoma japonicum"
                     /strain="Chinese (Anhui) strain"
                     /db_xref="taxon:6182"
                     /clone_lib="Adult SjC 7/94"
                     /sex="Male and female"
                     /tissue_type="Whole body"
                     /dev_stage="Adult worms"
                     /lab_host="Mouse and rabbit"
                     /note="Vector: Lambda ZAP-II XR.; Site_1: EcoR I; Site_2:
                     Xhol I; Several hundred adult Schistosoma japonicum
                     (Anhui, P.R. China, strain), of mixed sex, were perfused
                     from the mesenteries of experimentally infected mice and
                     rabbits at the Queensland Institute of Medical Research,
                     Brisbane, Australia (QIMR), and stored for several months
                     in liquid nitrogen. Subsequently, mRNA was isolated at the
                     QIMR from lysates of these worms by oligo dT
                     chromatography, using a kit from Pharmacia. The mRNA was
                     then shipped to Clontech, Palo Alto, CA, USA, who
                     constructed a cDNA library. First strand synthesis was
                     primed with an oligo-dT-XhoI-primer and synthesized using
                     M-MLV reverse transcriptase. Second strand synthesis was
                     accomplished with RNAse H and T4 DNA polymerase. The
                     double stranded cDNA was ligated to EcoRI linkers,
                     digested with EcoRI and XhoI, and ligated into the
                     phagemid vector lambda ZAP II XR. After construction of
                     this directional library by Clontech, it was returned to
                     the QIMR. During analysis of the library at the QIMR, we
                     have found that a small percentage, 2% to 3%, of the
                     clones contain inserts that appear to be highly homologous
                     to sequences from salmonoid fishes, as determined by
                     homology comparisons using BLAST and by Southern
                     hybridization analysis to genomic DNA from salmon (Sigma
                     Chemical Co., St. Louis, MO) under stringent washing
                     conditions. The remainder of the clones appear to contain
                     S. japonicum sequences."
BASE COUNT      110 a     89 c     85 g    115 t
ORIGIN


alignment_scores:
            Quality:   79.00            Length:    117
              Ratio:  1.039              Gaps:      6
 Percent Similarity:  64.957   Percent Identity: 25.641

alignment_block:
 US-09-068-804-6 x AI083069   ..

 Align seg 1/1  to: AI083069  from: 1  to: 399

      43 SerSerLysValSerAspLeuGlyLeuSerProThrLeuSerAlaProAl 59
         |||   :::|||||||:::   ||||||:::   ::::::::::::|||::
      49 TCCGTACGAGTCTCGGCATATGGCCTCGGCCTTATATTTGATGTTCCTAC 98

      59 a......ProGlyValLeuThrGlnThrProGlyThrIle..... 70
         :      :::   ||||||:::|||   :::|||:::|||
      99 GGGGAATGCAAAGGTATTATCCCAACACTCGGGCAGTATTCGAGTCAATG 148

      71 .....ThrSerPheLeuLysAlaSerIleGlnAsnThrAspMetAsnGln 85
               :::      |||::::::      |||:::|||:::   :::|||
     149 GACATGTTCATTGGCTTACGGATGAGTCTCAAAGTACCTCGAGTTCGCAG 198

      86 AspLeuAsnAlaLeuAlaAsnAsnValThrThrLysAlaAsnGluValVa 102
         :::|||:::::::|||:::::::::            :::   |||:::
     199 TCCCTTGACTCACTATCTGCAAGC............GATCCAGAGATCTC 236

     102 lGlnThrGlnLeuArgGlu............GlnGlnAlaGluValGlyL 115
          |||   |||   |||:::            |||:::::::::   ::::
     237 TCAGCATCAGCACAGAAATTTGAGAATACGCCAAGAAAGTGACAAAAGCA 286

     115 ys.........PhePheAspIleSerGlyMetSerSerSerAlaValAla 128
         ::         :::|||:::::::::::::::::||||||   :::::::::
     287 CTCACACTTCGTTGTTTAATGTACCAGATTTGTCTTCTGAAACAATTGGG 336

     129 LeuLeuAlaAlaAla.................AsnThrLeuMetLeuTh 139
         :::|||   ||||||                :::||||||:::::::
     337 TTTTTGAAAGCTGCTAGATCGTCATTGCCTCCTACTACTTTGCTTGTTGC 386

     139 r 139
         :
     387 A 387

seq_name: gb_est26:AI372258

seq_documentation_block:
LOCUS       AI372258      416 bp    mRNA            EST       12-JAN-1999
DEFINITION  au03g07.x1 maize ear glume library Zea mays cDNA clone au03g07 5',
            mRNA sequence.
ACCESSION   AI372258
NID         g4152124
VERSION     AI372258.1  GI:4152124
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Zea.
REFERENCE   1  (bases 1 to 416)
  AUTHORS   Norris,A., Parnell,L.D., Schutz,K., de la Bastide,M.,
            Nascimento,L., Shah,R., Swaby,I., Shekher,M., Vil,M.D., Preston,R.,
            Huang,E.N., Habermann,K., Dedhia,N., Martienssen,R. and
            McCombie,W.R.
  TITLE     Expressed sequence tags from Z. mays (1999)
  JOURNAL   Unpublished (1999)
COMMENT     On Apr 3, 1998 this sequence version replaced gi:3018011.

            Contact: W. Richard McCombie
            Lita Annenberg Hazen Genome Sequencing Center
            Cold Spring Harbor Laboratory
            PO Box 100, Cold Spring Harbor, NY 11724, USA
            Tel: 516 367 8884
            Fax: 516 367 8874
            Email: mccombie@cshl.org
            Plate: au03  row: g  column: 07
            Seq primer: M13 reverse universal
            High quality sequence stop: 416.
FEATURES             Location/Qualifiers
     source          1. .416
                     /organism="Zea mays"
                     /db_xref="taxon:4577"
                     /map="21q"
                     /clone="au03g07"
                     /clone_lib="maize ear glume library"
                     /tissue_type="glume ear 2 weeks post-pollination"
                     /cell_line="W22-TGA"
                     /note="Vector: pBLUESCRIPT SK+ (X52325); Site_1: XhoI;
                     Site_2: EcoRI; This library was supplied by John Doebley
                     (Univ. of Minnesota). cDNAs are  directionally cloned into
                     the XhoI and EcoRI sites; XhoI is near the polyA  tail.
                     Most reads from this library are made with the reverse
                     universal  M13 primer and are 5' in direction. Additional
                     information on this  library as well as ftp access to all
                     sequences can be found at
                     http://www.cshl.org/maizegenome"
BASE COUNT      133 a     96 c    116 g     71 t
```

```
     159 laLysThrThrAlaSerSerMetMetArgGluGlyMetAsnAlaLeuSer 175
             :::::|||||||      ||||||      :::
     333 GGGAAGGATCTGTATCAAGCCGTTCAAGAGAATGCAGTCCTGTCCCCATG 382

     176 GlySerIleSerGlnSerAlaLeuGlnLeuGlyIleThrGlyValGlyAl 192
         ||||||:::::::::   |||:::     |||                ::
     383 GGGTCATTCCCAAGAAGAGCATTTGTTAATGGAAGCAGAGAAAGCACTGG 432

     192 aLysLeuGluTyrLysGlyLeuGlnAsnGluArgGlyAlaLeu 206
         :   ||||||        |||::::::||||||:::   |||
     433 ATACTTAGAA......GAGCTTGAGAAAGAGAGATCATTACTC 469

seq_name: gb_est17:AA624987

seq_documentation_block:
LOCUS       AA624987      461 bp    mRNA            EST       14-OCT-1997
DEFINITION  JE00F100.QF3 PME-SD.96 Schistosoma japonicum cDNA clone SJPF100 5'
            similar to Fibrillin 2, mRNA sequence.
ACCESSION   AA624987
NID         g2528863
VERSION     AA624987.1  GI:2528863
KEYWORDS    EST.
SOURCE      Schistosoma japonicum.
  ORGANISM  Schistosoma japonicum
            Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
            Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.
REFERENCE   1  (bases 1 to 461)
  AUTHORS   Fan,J. and Brindley,P.J.
  TITLE     Expressed sequence tags (ESTs) from a cDNA library of Schistosoma
            japonicum (Philippine strain) Egg
  JOURNAL   Unpublished (1997)
COMMENT     On May 9, 1995 this sequence version replaced gi:802330.

            Contact: Brindley, P.J.
            Molecular Parasitology Unit
            Queensland Institute of Medical Research
            300 Herston Road, Qeensland 4029, Australia
            Tel: 61 7 3362 0413
            Fax: 61 7 3362 0104
            Email: paulB@qimr.edu.au
            PCR PRimers
            FORWARD: M13 Forward
            BACKWARD: M13 Reverse
            Insert Length: 1200   Std Error: 0.00
            Seq primer: T3 Reverse
            High quality sequence stop: 461.
FEATURES             Location/Qualifiers
     source          1. .461
                     /organism="Schistosoma japonicum"
                     /strain="Philippine strain"
                     /db_xref="taxon:6182"
                     /map="15"
                     /clone="SJPF100"
                     /clone_lib="PME-SD.96"
                     /sex="Mixed sex"
                     /tissue_type="Whole organism"
                     /dev_stage="Egg"
                     /lab_host="Rabbit"
                     /note="Vector: Lambda ZAP-II XR (Stratagene); Site_1:
                     EcoRI; Site_2: Xho I; Eggs of Schistosoma japonicum
                     (Philippine strain) were purified from livers of rabbits
                     about 50 days after infection with cercariae at the
                     College of Public Health, University of The Philippines,
                     Manila, The Philippines. Subsequently, at The Queensland
                     Institute of Medical Research, Brisbane, Australia,
                     messenger RNA was isolated from lysates of the S.
                     japonicum eggs by oligo dT chromatography, first strand
                     cDNA synthesis from the mRNA was primed with an oligo
                     dT-XhoI primer and synthesized with M-MLV reverse
                     transcriptase, and second strand synthesis was
                     accomplished with RNAse H and T4 DNA polymerase. The
                     double stranded cDNA was ligated into the phagemid vector
                     UNI ZAP XR (Stratagene), and the ligation products
                     packaged using reagents from Stratagene. The library was
                     amplified using BB4 strain E. coli bacteria. The library
                     was constructed by Sharon Day, Rolando Tan, Wilfred Tiu,
                     and Paul Brindley who consider that it contains sequences
                     expressed by immature and mature eggs of S. japonicum. The
                     library is named 'PME-SD.96'"
BASE COUNT      128 a    106 c     96 g    131 t
ORIGIN


alignment_scores:
           Quality:   80.00            Length:    103
             Ratio:  1.194              Gaps:      5
 Percent Similarity:  65.049   Percent Identity: 26.214

alignment_block:
 US-09-068-804-6 x AA624987   ..

 Align seg 1/1  to: AA624987  from: 1  to: 461

      47 SerAspLeuGlyLeuSerProThrLeu.....SerAlaProAl 59
         ::::::||||||:::   |||   |||          ::::::   ::
      46 ACGAGTCTCGGCATATGGCCTCGGCTTATATTTGATGTTCCTACGGGGAA 95

      59 aProGlyValLeuThrGlnThrProGlyThrIle.............T 71
         ::::   ||||||:::|||   :::|||:::|||             :
      96 TGCAAAGGTATTATCCCAACACTCGGGCAGTATTCGAGTCAATGGACATG 145

      71 hrSerPheLeuLysAlaSerIleGlnAsnThrAspMetAsnGlnAspLeu 87
         ::      |||::::::      |||:::|||:::   :::|||:::|||
     146 TTCATTGGCTTACGGATGAGTCTCAAAGTACCTCGAGTTCGCAGTCCCTT 195

      88 AsnAlaLeuAlaAsnAsnValThrThrLysAlaAsnGluValValGlnTh 104
         ::::::|||:::::::::            :::   |||:::   |||
     196 GACTCACTATCTGCAAGC.....GATCCAGAGATCTCTCAGCA 233

     104 rGlnLeuArgGlu.....GlnGlnAlaGluValGlyLys.... 115
          |||||||||:::          |||:::::::::   ::::::
     234 TCAGCTCAGAAATTTGAGAATACGCCAAGAAAGTGACAAAAGCACTCACA 283

     116 .....PhePheAspIleSerGlyMetSerSerSerAlaValAlaLeuLeu 130
             :::|||:::::::::::::::::::||||||   ::::::::::::|||
     284 CTTCGTTGTTTAATGTACCAGATTTGTCTTCTGAAACAATTGGGTTTTTG 333

     131 AlaAlaAla 133
         ||||||
     334 AAAGCTGCT 342

seq_name: gb_est23:AI083069

seq_documentation_block:
LOCUS       AI083069      399 bp    mRNA            EST       17-AUG-1998
DEFINITION  JAA00068.QA3 Adult SjC 7/94 Schistosoma japonicum cDNA 5' similar
            to fibrillin 2, mRNA sequence.
ACCESSION   AI083069
NID         g3421492
VERSION     AI083069.1  GI:3421492
KEYWORDS    EST.
SOURCE      Schistosoma japonicum.
  ORGANISM  Schistosoma japonicum
            Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
            Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.
REFERENCE   1  (bases 1 to 399)
  AUTHORS   Gao,X., Brindley,P.J. and Feng,Z.
  TITLE     Expressed sequence tags from Adults of Schistosoma japonicum (Anhui
            strain)
  JOURNAL   Unpublished (1998)
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2150309.

            Contact: Brindley, P.J.
            Molecular Parasitology Unit
            Queensland Institute of Medical Research
```

```
alignment_scores:
            Quality:   85.00              Length:    144
              Ratio:  0.904                Gaps:      5
 Percent Similarity:  65.278   Percent Identity: 21.528

alignment_block:
 US-09-068-804-6 x AA618975   ..

 Align seg 1/1  to: AA618975  from: 1  to: 459

      43 SerSerLysValSerAspLeuGlyLeuSerProThrLeuSerAlaProAl 59
         |||   :::|||||||:::   ||||||:::   :::::::::::::|||:: 
      33 TCCGTACGAGTCTCGGCATATGGCCTCGGCCTTATATTTGATGTTCCTAC 82

      59 a......ProGlyValLeuThrGlnThrProGlyThrIle..... 70
         :      :::   ||||||:::|||   :::|||:::||| 
      83 GGGGAATGCAAAGGTATTATCCCAACACTCGGGCAGTATTCGAGTCAATG 132

      71 .....ThrSerPheLeuLysAlaSerIleGlnAsnThrAspMetAsnGln 85
            :::      |||:::::      |||:::|||:::   :::||| 
     133 GACATGTTCATTGGCTTACGGATGAGTCTCAAAGTACCTCGAGTTCGCAG 182

      86 AspLeuAsnAlaLeuAlaAsnAsnValThrThrLysAlaAsnGluValVa 102
         :::|||:::::::|||:::::::::            :::   |||::: 
     183 TCCCTTGACTCACTATCTGCAAGC.............GATCCAGAGATCTC 220

     102 lGlnThrGlnLeuArgGlu.............GlnGlnAlaGluValGlyL 115
          |||   |||   |||:::            |||:::::::::   :::: 
     221 TCAGCATCAGCACAGAAATTTGAGAATACGCCAAGAAAGTGACAAAAGCA 270

     115 ys.........PhePheAspIleSerGlyMetSerSerSerAlaValAla 128
         ::        :::|||:::::::::::::::::::|||||||   ::::::::: 
     271 CTCACACTTCGTTGTTTAATGTACCAGATTTGTCTTCTGAAACAATTGGG 320

     129 LeuLeuAlaAlaAlaAsnThrLeuMetLeuThrLeuAsnGlnAlaAspSe 145
         :::|||   ||||||   :::   :::   :::   :::          :: 
     321 TTTTTGAAAGCTGCTAGATCGTCATTACCTCCTACTACTTTGCTTGTTGC 370

     145 rLysLeuSerGlyLysLeuSerLeuValSerPheAspAlaAlaLysThrT 162
         ::::::::::::::   :::   |||   |||   ::::::   :::: 
     371 ACGTATTGGCGATACAAAGACAGCAGTCATATTTTTGTCCGTAGCTTCAA 420

     162 hrAlaSerSerMetMetArgGluGlyMetAsn 172
         ::::::::   :::::::::      |||::: 
     421 AACCTCCAGTGGTCCTTCAGAAAGAAATGCAC 452

seq_name: gb_est27:AA955065

seq_documentation_block:
LOCUS       AA955065      485 bp    mRNA            EST       07-FEB-1999
DEFINITION  UI-R-A1-ey-g-09-0-UI.s1 UI-R-A1 Rattus norvegicus cDNA clone
            UI-R-A1-ey-g-09-0-UI 3' similar to gi|928855|dbj|D38629|RATAPC Rat
            mRNA for APC protein, complete cds, mRNA sequence.
ACCESSION   AA955065
NID         g4237894
VERSION     AA955065.1  GI:4237894
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 485)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     On May 7, 1998 this sequence version replaced gi:3119114.

            Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            Oligo-dT track not found, Not I site shown in beginning of sequence
            is likely internal to the message. cDNA Library Preparation: M.
            Fatima Bonaldo, Ph.D. Clone distribution: clones will be available
            through Research Genetics
            Seq primer: M13 Forward.
FEATURES             Location/Qualifiers
     source          1. .485
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-A1-ey-g-09-0-UI"
                     /clone_lib="UI-R-A1"
                     /dev_stage="adult"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-A1
                     library is a subtracted library derived from the UI-R-A0
                     library. The UI-R-A0 library consisted of a mixture of
                     individually tagged normalized libraries constructed from
                     rat placenta, adult lung, brain, liver, kidney, heart,
                     spleen, ovary, and muscle. The tag is a string of 3-5
                     nucleotides present between the Not I site and the
                     oligo-dT track which allows identification of the library
                     of origin of a clone within the mixture. The subtracted
                     library (UI-R-A1) was constructed as follows: PCR
                     amplified cDNA inserts from a pool of approximately 3,840
                     UI-R-A0 clones from which 3' ESTs had been derived was
                     used as a driver in a hybridization with the UI-R-A0
                     library in the form of single-stranded circles. The
                     remaining single-stranded circles (subtracted library) was
                     purified by hydroxyapatite column chromatography,
                     converted to double-stranded circles and electroporated
                     into DH10B bacteria (Life Technologies) to generate the
                     UI-R-A1 library. This procedure has been previously
                     described (Bonaldo, Lennon and Soares, Genome Research 6:
                     791-806, 1996)"
BASE COUNT      153 a    100 c    124 g    108 t
ORIGIN


alignment_scores:
            Quality:   83.00              Length:    131
              Ratio:  1.038                Gaps:      3
 Percent Similarity:  61.069   Percent Identity: 25.954

alignment_block:
 US-09-068-804-6 x AA955065   ..

 Align seg 1/1  to: AA955065  from: 1  to: 485

      78 IleGlnAsnThrAspMetAsnGlnAspLeuAsnAlaLeuAlaAsnAsnVa 94
         :::::::|||:::::::::   |||:::|||::::::   :::|||:::::
      95 ATGGAGAACTCAAATCTTCGACAAGAGCTCGAAGATAATTCCAATCACCT 144

      94 lThrThrLysAlaAsnGluVal......ValGlnThrGlnLeuArgGluG 109
         :|||      :::::::|||:::      ::::::      |||:::::: 
     145 TACAGAACTGGAAACTGAGGCATCTAACATGAAGGAAGTACTTAAGCAGC 194

     109 lnGlnAlaGluValGlyLysPhePheAspIleSerGlyMetSerSerSer 125
          |||:::   :::            :::         |||:::|||::: 
     195 TGCAGGGAAGTATT.............GAAGATGAGACTATGACTTCCGGA 232

     126 AlaValAlaLeuLeuAlaAlaAlaAsnThrLeuMetLeuThrLeuAsnGl 142
         ::::::||||||:::   :::   :::  |||      |||
     233 CAGATTGATTTACTAGAGCGCCTTAAAGAATTTAACTTAGATAGCAATTT 282

     142 nAlaAspSerLysLeuSerGlyLysLeuSerLeuValSerPheAspAlaA 159
         ::::::   ||||||   :::|||:::||||||   |||:::::::::
     283 CCCCGGAGTGAAACTACGCTCGAAAATGTCCCTCCGATCCTATGGAAGTC 332
```

OM of: US-09-068-804-6 to: EST:* out_format : pfs

Date: Oct 5, 1999 7:34 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09068804/runat_05101999_145134_10306/app_query.fasta.1
-DB=EST -QFMT=fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=pam150 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=escore -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfs -NORM=stat -USER=US09068804 -NCPU=6 -ICPU=3 -WAIT
-THREADS=1

Search information block:
Query: US-09-068-804-6
Query length: 410
Database: EST:*
Database sequences: 2546578
Database length: 986266752
Search time (sec): 940.790000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ..
gb_est17:AA618975	+	85.00	186.78	0.0688	459	! AA618975 JE00F077.QF3 PME-SD.96
gb_est27:AA955065	+	83.00	181.31	0.1388	485	! AA955065 UI-R-A1-ey-g-09-0-UI.s
gb_est17:AA624987	+	80.00	174.64	0.3262	461	! AA624987 JE00F100.QF3 PME-SD.96
gb_est23:AI083069	+	79.00	173.90	0.3587	399	! AI083069 JAA00068.QA3 Adult SjC
gb_est26:AI372258	+	79.00	173.42	0.3817	416	! AI372258 au03g07.x1 maize ear g
gb_est25:AU039285	+	80.50	171.99	0.4585	643	! AU039285 AU039285 Dictyostelium
gb_est24:AU034008	+	80.50	171.20	0.5070	688	! AU034008 AU034008 Dictyostelium
gb_est25:AI294100	-	79.00	170.72	0.5397	525	! AI294100 LP07402.5prime LP Dros
gb_est16:AA560346	-	78.00	170.54	0.5524	433	! AA560346 vl39a10.r1 Stratagene
gb_est25:AI295460	-	79.00	170.48	0.5567	536	! AI295460 LP09102.5prime LP Dros
gb_est17:C63675	+	77.00	170.26	0.5722	360	! C63675 C63675 Yuji Kohara unpub
gb_est28:AI516432	-	79.00	170.24	0.5738	547	! AI516432 LD42296.5prime LD Dros
gb_est19:AA760238	-	77.00	169.82	0.6056	374	! AA760238 vv70h07.r1 Stratagene
gb_est14:AA391538	-	78.00	169.49	0.6321	474	! AA391538 LD10548.5prime LD Dros
gb_est3:H12259	+	78.50	169.44	0.6356	528	! H12259 ym17d09.r1 Soares infant
gb_est16:AA563507	+	80.00	168.30	0.7356	796	! AA563507 JE00F001.QF3 PME-SD.96
gb_est10:AA167844	+	78.50	165.24	1.09	758	! AA167844 JM00M085.QM3 Miracidia
gb_est25:AU037986	+	77.50	165.03	1.12	627	! AU037986 AU037986 Dictyostelium
gb_est19:AA771437	+	77.50	164.99	1.12	629	! AA771437 vm45a04.r1 Stratagene
gb_est10:AA161028	-	76.50	164.22	1.24	546	! AA161028 zo58d07.s1 Stratagene
gb_est8:AA019153	+	72.50	162.72	1.50	270	! AA019153 ze56e07.r1 Soares reti
gb_est21:C90107	+	76.50	162.32	1.58	643	! C90107 C90107 Dictyostelium dis
gb_est21:C90237	+	76.00	161.86	1.68	603	! C90237 C90237 Dictyostelium dis
gb_est19:AA392341	-	74.00	160.06	2.12	464	! AA392341 LD11181.5prime LD Dros
gb_est21:AA950134	+	76.00	159.96	2.14	710	! AA950134 LD28672.5prime LD Dros
gb_est11:AA264874	-	75.50	159.58	2.25	661	! AA264874 LD09545.5prime LD Dros
gb_est28:AI530269	+	75.50	158.54	2.57	723	! AI530269 ui91a11.y1 Sugano mous
gb_est25:AI316613	+	75.00	157.67	2.88	702	! AI316613 uk62c09.y1 Schiller mo
gb_est28:AI529737	+	76.00	157.35	3.00	889	! AI529737 ui81g10.y1 Sugano mous
gb_est27:AI398365	-	72.50	156.95	3.16	444	! AI398365 NCSC1G7T7 Subtracted C
gb_est31:AI691735	+	73.50	156.57	3.31	565	! AI691735 606007F10.x1 606 - Ear
gb_est27:AI425728	+	74.00	156.04	3.54	656	! AI425728 me69f09.y1 Soares mous
gb_est16:AA550027	-	73.50	156.01	3.56	593	! AA550027 l106m3 qmbPfHB3.1, G.
gb_est3:R63488	+	72.50	155.95	3.59	484	! R63488 yi08g10.r1 Soares placen
gb_est8:AA029002	+	72.50	155.95	3.59	484	! AA029002 zk11e08.r1 Soares_preg
gb_est16:C42963	+	71.00	155.76	3.68	360	! C42963 C42963 Yuji Kohara unpub
gb_est28:AI484415	+	73.00	155.20	3.95	573	! AI484415 EST242425 tomato ovary
gb_est15:AA477043	+	71.50	155.18	3.96	420	! AA477043 zu38d05.r1 Soares ovar
gb_est16:AA537947	+	73.00	154.90	4.11	588	! AA537947 vj36b09.r1 Stratagene
gb_est17:AA606469	+	73.00	154.26	4.45	621	! AA606469 vo42g12.r1 Barstead mo
gb_est28:AI526540	-	74.00	154.14	4.53	773	! AI526540 uj39f11.y1 Sugano mous
gb_est28:AI526630	-	73.50	154.13	4.53	697	! AI526630 uj40h12.y1 Sugano mous
gb_est8:AA032396	+	73.00	154.09	4.55	630	! AA032396 mi33b08.r1 Soares m
gb_est26:AI388899	+	72.00	153.32	5.03	547	! AI388899 GH19957.5prime GH D
gb_est22:AI047411	+	71.50	153.27	5.06	495	! AI047411 ud65h12.y1 Sugano m

seq_name: gb_est17:AA618975

seq_documentation_block:
LOCUS AA618975 459 bp mRNA EST 09-OCT-1997
DEFINITION JE00F077.QF3 PME-SD.96 Schistosoma japonicum cDNA clone SJPF077 5' similar to Fibrillin 2, mRNA sequence.
ACCESSION AA618975
NID g2522851
VERSION AA618975.1 GI:2522851
KEYWORDS EST.
SOURCE Schistosoma japonicum.
ORGANISM Schistosoma japonicum
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.
REFERENCE 1 (bases 1 to 459)
AUTHORS Fan,J. and Brindley,P.J.
TITLE Expressed sequence tags (ESTs) from a cDNA library of Schistosoma japonicum (Philippine strain) Egg
JOURNAL Unpublished (1997)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1404944.

Contact: Brindley, P.J.
Molecular Parasitology Unit
Queensland Institute of Medical Research
300 Herston Road, Qeensland 4029, Australia
Tel: 61 7 3362 0413
Fax: 61 7 3362 0104
Email: paulB@qimr.edu.au
PCR PRimers
FORWARD: M13 Forward
BACKWARD: M13 Reverse
Insert Length: 1100 Std Error: 0.00
Seq primer: T3 Reverse
High quality sequence stop: 459.

FEATURES Location/Qualifiers
source 1. .459
/organism="Schistosoma japonicum"
/strain="Philippine strain"
/db_xref="taxon:6182"
/clone="SJPF077"
/clone_lib="PME-SD.96"
/sex="Mixed sex"
/tissue_type="Whole organism"
/dev_stage="Egg"
/lab_host="Rabbit"
/note="Vector: Lambda ZAP-II XR (Stratagene); Site_1: EcoRI; Site_2: Xho I; Eggs of Schistosoma japonicum (Philippine strain) were purified from livers of rabbits about 50 days after infection with cercariae at the College of Public Health, University of The Philippines, Manila, The Philippines. Subsequently, at The Queensland Institute of Medical Research, Brisbane, Australia, messenger RNA was isolated from lysates of the S. japonicum eggs by oligo dT chromatography, first strand cDNA synthesis from the mRNA was primed with an oligo dT-XhoI primer and synthesized with M-MLV reverse transcriptase, and second strand synthesis was accomplished with RNAse H and T4 DNA polymerase. The double stranded cDNA was ligated into the phagemid vector UNI ZAP XR (Stratagene), and the ligation products packaged using reagents from Stratagene. The library was amplified using BB4 strain E. coli bacteria. The library was constructed by Sharon Day, Rolando Tan, Wilfred Tiu, and Paul Brindley who consider that it contains sequences expressed by immature and mature eggs of S. japonicum. The library is named 'PME-SD.96'"

BASE COUNT 126 a 106 c 94 g 133 t
ORIGIN

This Page Blank (uspto)

```
  JOURNAL   Submitted (29-DEC-1998) segalat L.S., Institut de Pharmacologie
            Moleculaire et Cellulaire, CNRS, 660 route des lucioles, Valbonne,
            06560, FRANCE
REFERENCE   2  (bases 1 to 1773)
  AUTHORS   Gieseler,K., Bessou,C. and Segalat,L.
  TITLE     Dystrobrevin- and dystrophin-like mutants display similar
            phenotypes in the nematode C. elegans
  JOURNAL   Unpublished
FEATURES             Location/Qualifiers
     source          1. .1773
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
     CDS             1. .1773
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAA10498.1"
                     /db_xref="PID:e1374530"
                     /db_xref="PID:g4218165"
                     /db_xref="GI:4218165"
                     /translation="MLWSNGGGGPREPSSAPSPDHHRAMHSVPPIVASEMQQLIDEMR
                     LQDFDSIRFATYRAACKLRFIQQKTKVHLVDIWNMIEAFRENGLNALPLHTVIKTSRA
                     ELLLTTVFHNLNKRLVASQHVDTDVSISLLLAFLLGAYDKQNTGRLTVFSIKVALATL
                     CAGKLVDKLRYIFSQIADSNGLMDHIKFTDFLQQILSLTTAVFEAPTFGFSENAVNQC
                     FHKDEKVSLNVFLDTFLSDPCPPCIMWLPLLHRMASVSNVYHPVVCDACQVRSFTGFR
                     YKCQRCANYQLCQSCFWRGRTSQNHSNEHEMKEYSSYKSPTKQLVHSIHKSLQCIPAT
                     STVGDANIDIFNAKIGGPVSSKPARPLNLNNIVPATPTTIRRQHAATSSADWPTSPVL
                     LPGQASHGGVIDDEHKLIARYAAKLSGRADYPLSNGRSMNSSMVGDERTLIAQLEEEN
                     SMMVREMARLESQTTSDDGLAGLRDRKMELEEKMFEMQQRRRELMMQLEHLMAQLNTG
                     PQPSGGVSSASLSQLPFTASDQQLTGVNSTVANAFRAGSLPATSLQGDLLHAADQITT
                     NMSDLVRQLDLAQSDENGVTINGF"
     gene            112
                     /gene="dyb-1"
     allele          112
                     /gene="dyb-1"
                     /note="allele cx36 (EMS-induced)"
                     /replace="t"
BASE COUNT      512 a    394 c    414 g    453 t
ORIGIN


alignment_scores:
            Quality:   84.00              Length:    234
              Ratio:  0.609                Gaps:      8
 Percent Similarity:  58.974   Percent Identity: 18.803

alignment_block:
 US-09-068-804-7 x CEL131742   ..

 Align seg 1/1  to: CEL131742  from: 1  to: 1773

      91 AlaLysSerGlyValSerLeuSerAlaGluGlnAsnGluAsnLeuArgSe 107
         :::::::::::::   :::   |||:::|||::::::   :::      ||
     883 GGAAGGACGAGTCAAAATCATTCAAATGAGCACGAGATGAAAGAATATTC 932

     107 rAlaPheSerAlaProThrSerAlaLeuPheSerAlaSerProMetAla. 123
         |::::::   :::||||||      |||      :::      ::::::
     933 CAGCTATAAATCGCCGACAAAACAGCTCGTACACTCAATTCACAAAAGTC 982

     124 ...........GlnProArgThrThrIleSerAspAlaGluIle...... 134
                    ::::::   :::|||:::::::|||||||:::|||
     983 TCCAATGTATTCCTGCCACGTCAACCGTCGGCGATGCAAACATCGATATT 1032

     135 ...TrpAspMetValSerGlnAsnIleSerAlaIleGlyAspSerTyrLe 150
              :::::::::::::      :::|||:::      :::      ||
    1033 TTTAACGCAAAAATCGGTGGCCCGGTCAGCAGCAAACCTGCTCGACCGCT 1082

     150 uGlyValTyrGluAsnValValAlaValTyrThrAspPheTyrGlnAlaP 167
         |::::::   :::   |||   |||:::  |||   :::  :::
    1083 AAACCTGAACAACATCGTCCCGGCGACACCGACGACGATCCGCCGGCAAC 1132

     167 heSerAspIleLeuSerLysMetGlyGlyTrp..............Leu 178
           :::::::::   |||      ::::::|||              |||
    1133 ACGCTGCCACGTCATCA......GCGGACTGGCCGACGTCACCGGTACTT 1176

     179 LeuProGlyLysAspGlyAsnThrValLysLeuAspValThrSerLeuLy 195
         |||||||||:::::::::::                        ::::::
    1177 CTGCCGGGACAGGCGTCGCAT.................GGAGGAGTTAT 1208

     195 sAsnAspLeuAsnSerLeuValAsnLysTyrAsnGlnIleAsnSerAsnT 212
          :::|||   :::   |||:::::::::::|||:::         |||:::
    1209 TGATGATGAGCATAAGCTTATTGCGAGATATGCAGCAAAGTTGTCGGGAA 1258

     212 hrValLeuPheProAlaGlnSerGlySerGlyValLysValAlaThrGlu 228
           :::   :::|||      :::|||   :::::::::   :::
    1259 GGGCTGATTATCCCCTATCAAACGGTCGAAGTATGAATAGTTCAATGGTC 1308

     229 AlaGluAlaArgGlnTrpLeuSerGluLeuAsnLeuProAsnSerCysLe 245
         :::::::::|||      ::::::::::|||:::      ||||||
    1309 GGAGACGAACGGACCCTTATTGCTCAACTGGAAGAAGAAAATTCA..... 1353

     245 uLysSerTyrGlySerGlyTyrValValThrValAspLeuThrProLeuG 262
                           :::::::::   :::::::::   |||:
    1354 .....................ATGATGGTACGGGAAATGGCTCGACTGG 1381

     262 lnLysMetValGlnAspIleAspGlyLeuGlyAlaProGlyLysAspSer 278
         ::      :::   :::   |||||||||:::::      :::|||
    1382 AGAGCCAAACAACTTCCGATGACGGGCTCGCCGGGCTC...CGCGACCGG 1428

     279 LysLeuGluMetAspAsnAlaLysTyr......GlnAlaTrpGlnSerGl 293
         |||:::|||:::::::::   ::::::      |||   ::::::
    1429 AAAATGGAATTGGAAGAGAAAATGTTTGAAATGCAACAGAGGAGGAGAGA 1478

     293 yPheLysAlaGlnGluGluAsnMetLysThrThrLeuGlnThrLeuThrG 310
          ::::::   |||   |||:::::::::::::   |||:::|||   :::|
    1479 ATTAATGATGCAATTGGAGCATTTAATGGCTCAGTTGAATACGGGACCCC 1528

     310 ln 310
         ||
    1529 AA 1530
```

```
                     ISGRPGQIFVQRRLASPWCSRCWRRTDCRGRSTPITIGLLLKGERRHGEGERSSNIFA
                     RASAAFERGFEHLRDGYSDLLLTLLRRRVIVPVVAVLVLALGATMFVYVGRDFYPLID
                     GGQIQLHVRAPAGTRIERTEAIFQAVEDKIREVIPKRDRALIVDNIGLPARPYNLAFT
                     DGSTIGVNDGTILVSLKDGHKPTADYIRKLRQVLPSAFPEDTFYFQAADIVTQILNFG
                     LPAQIDVRTVGYGSNNLAVAKELQKSLAAIPGVVDAHLQQEVDAPAFYADIDRTRAAQ
                     LGLNASTVATNINVSLSSSAQVSPNFWTDPTSGIPYYLAVQTPEYRANSLNSLGNTPV
                     SASLAASGQTVPGLLSNVATFKRGTVPTNLNQANVQPVFDVYASVQGRDLGSVASDID
                     KVTSTLQKQLQPGNSIQVLGQIQSMHQSFRDLGIGLLFAAILVYLLMVVNYQNFGDPF
                     VVIMALPATFCGIVTMLFVTGTTLNVPSLMGAIMAVGVASANSILLVTFARDEQLKGH
                     SAFQAALSAGRTRIRPVLKTAAAMIVGMIPMAIGGPGEEQNAALARAVIGGLLFATPT
                     TLLIVPYLFAMLRKGNDGKPRHGVFEEQPE"
     gene            3240. .4511
                     /gene="ragD"
     CDS             3240. .4511
                     /gene="ragD"
                     /function="membrane fusion protein"
                     /codon_start=1
                     /transl_table=11
                     /product="RagD"
                     /protein_id="CAA12354.1"
                     /db_xref="PID:e1270776"
                     /db_xref="PID:g2961294"
                     /db_xref="GI:2961294"
                     /db_xref="SPTREMBL:O69239"
                     /translation="MSDVRERSDDDEAADRPGAESLRIVELPGKVARSKRRYGGPLLG
                     GGVLLLLAGGLGIGGWRHYEAAREVAAVAEQVRTNVPEVRVATVQASGDLMKVTLPAT
                     TTAFEAANIFARTSGYIEKRYVDIGDRVKKGDLLVEITAPELDQQIAQAQATLAQDQA
                     ALQQAQASRELADVTNSRDSNLVKQGWLTAQQGDNDRLTLRAQQGAVGVAQSNIAAQE
                     AQIRVLQQEKSYQRVVAPFDGVVTQRNVDNGSLVQAGSTFMFTMMHSNVIRTQVFVPQ
                     DEAFGVAPGVDSDIRVPEIPGRTFAGKVTRIATALQPGSRTLLTEIDVPNPDGLLSPG
                     IYCTVELSIPRRTPSTTIPSDALVFDQNGLHVAVVRNGTVHFQQVSISRDFGTTVEVR
                     DGVQPGDQVVLNPAVNLAEGSKVTVRKSEVS"
BASE COUNT     1070 a    1903 c    1821 g    1071 t
ORIGIN


alignment_scores:
            Quality:   90.50              Length:    338
              Ratio:  0.545                Gaps:     16
 Percent Similarity:  49.112   Percent Identity: 21.302

alignment_block:
 US-09-068-804-7 x BJA225023   ..

 Align seg 1/1  to: BJA225023  from: 1  to: 5865

       5 GlnAsnTyrSerAlaSerProHisProGlyIleValAlaGluArgProGl 21
         ::::::|||   ||:::   :::   :::   ||||||||||:::   ::
    3420 CGGCACTATGAGGCCGCGCGTGAGGTCGCCGCAGTGGCAGAGCAGGTTCG 3469

      21 nThrProSerAlaSerGluHisValGluThr..................... 31
         :|||      :::      :::|||:::|||
    3470 GACCAACGTTCCGGAAGTTCGCGTGGCGACGGTCCAGGCTAGCGGCGACC 3519

      32 ........AlaValValProSerThrThr...................... 38
                 :::::::::|||:::||||||
    3520 TCATGAAGGTCACCTTGCCGGCGACGACGACCGCGTTCGAGGCTGCCAAC 3569

      39 ............................GluHisArgGlyThrAspIle.. 45
                                     |||   |||   :::||||||
    3570 ATCTTTGCGCGGACCAGCGGCTATATCGAGAAGCGCTACGTCGACATCGG 3619

      46 ................................IleSerLeuSerGlnA 51
                                         |||:::   ::::::
    3620 CGATCGAGTGAAGAAGGGCGATCTCCTCGTGGAAATCACCGCTCCCGAGC 3669

      51 laAlaThrLysIleHisGlnAlaGlnGlnThrLeuGlnSerThrProPro 67
           :::   :::|||   |||||||||   ||||||         :::::
    3670 TGGACCAGCAGATCGCGCAGGCGCAGGCGACGCTCGCGCAGGACCAGGCC 3719

      68 IleSerGluGluAsnAsnAspGluArgThrLeuAlaArgGlnGlnLeuTh 84
           ::::::::::::::   |||   ||||||      :::::::||
    3720 GCACTTCAGCAAGCGCAAGCGAGCCGAGAACTGGCC......GACGTCAC 3763

      84 rSerSerLeuAsnAla...LeuAlaLysSerGlyValSerLeuSerAlaG 100
         |:::|||   ::::::   |||:::|||   |||      |||:::|||:
    3764 CAATTCGCGCGACAGCAATCTCGTCAAGCAGGGC...TGGTTGACGGCGC 3810

     100 luGlnAsnGluAsnLeuArgSerAlaPheSerAlaProThrSerAlaLeu 116
         ::|||::::::|||   |||   ::::::   |||:::   :::|||:::
    3811 AACAGGGCGATAACGATCGTCTTACCCTGCGCGCACAGCAGGGCGCGGTG 3860

     117 PheSerAlaSerProMetAlaGlnProArgThrThrIleSerAspAlaGl 133
                       :::||||||:::   ::::::      :::|||::
    3861 ............GGCGTCGCGCAGTCGAACATCGCGGCGCAAGAGGCGCA 3898

     133 uIleTrpAspMetValSerGlnAsnIleSerAlaIleGlyAspSerTyrL 150
         :|||:::   :::      ::::::
    3899 GATTCGCGTGCTCCAGCAGGAGAAG........................... 3923

     150 euGlyValTyrGluAsnValValAlaValTyrThrAspPheTyrGlnAla 166
               |||:::   |||||||||                      :::
    3924 .....TCCTACCAGCGCGTGGTGGCG.....................CCG 3947

     167 PheSerAspIleLeuSerLysMetGlyGlyTrpLeuLeuProGlyLysAs 183
         |||:::::::::
    3948 TTCGACGGCGTG..................................... 3959

     183 pGlyAsnThrValLysLeuAspValThrSerLeuLysAsnAspLeuAsnS 200
                                |||||      :::   |||   :::|
    3960 .....................GTCACGCAGCGCAACGTGGACAATGGCA 3987

     200 erLeuValAsn..............LysTyrAsnGlnIleAsnSerAsn 211
         ||||||||:::              :::::::::   ::::::||||||
    3988 GCCTAGTGCAGGCCGGCTCCACCTTCATGTTCACGATGATGCATTCCAAT 4037

     212 Thr............ValLeuPheProAlaGlnSerGlySerGlyValLy 224
         :::           |||:::   |||   :::   :::   ||||||
    4038 GTCATCCGAACCCAGGTCTTCGTGCCGCAGGACGAGGCTTTCGGCGTTGC 4087

     224 sValAlaThrGluAlaGluAlaArgGlnTrpLeuSerGluLeuAsnLeuP 241
             ::::::::::::::::   |||                    :::|
    4088 TCCGGGCGTCGACTCGGATATCCGC.....................GTTC 4116

     241 roAsnSerCysLeuLysSerTyrGlySerGlyTyrValValThrValAsp 257
         ||:::          ::::::::::::::::          ||||||
    4117 CCGAGATTCCGGGGCGGACGTTTGCGGGCAAG......GTCACGCGGATC 4160

     258 LeuThrProLeuGln............LysMetValGlnAspIle..... 268
             |||:::||||||            :::::::::   ::::||
    4161 GCGACCGCGCTGCAGCCGGGGAGCCGGACGTTGCTGACCGAGATCGACGT 4210

     269 ..........AspGlyLeuGlyAlaProGlyLysAspSerLysLeuGluM 282
                  |||||||||   :::||||||      ::::::::::|||:
    4211 GCCCAACCCCGATGGCCTGCTCAGTCCCGGCATCTATTGCACGGTCGAAT 4260

     282 etAspAsnAlaLys 286
         :::::   ::::::
    4261 TATCGATCCCGCGC 4274

seq_name: gb_in1:CEL131742

seq_documentation_block:
LOCUS       CEL131742    1773 bp    mRNA            INV       02-FEB-1999
DEFINITION  Caenorhabditis elegans mRNA for dystrobrevin.
ACCESSION   AJ131742
NID         g4218164
VERSION     AJ131742.1  GI:4218164
KEYWORDS    dyb-1 gene; dystrobrevin.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 1773)
  AUTHORS   segalat,L.S.
  TITLE     Direct Submission
```

```
  TITLE     Studies of manganese-superoxide dismutase gene of Ganoderma
  JOURNAL   Thesis (1996)
REFERENCE   2  (bases 1 to 742)
  AUTHORS   Wang,H-F. and Hseu,R.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-APR-1996) Ruey-Shyang Hseu, Agricultural Chemistry,
            Applied Microbiology, National Taiwan University, Taipei, Taiwan,
            Republic of China
FEATURES             Location/Qualifiers
     source          1. .742
                     /organism="Ganoderma tsugae"
                     /strain="RSH J2"
                     /db_xref="taxon:34467"
     gene            1. .742
                     /gene="Mn-SOD"
     CDS             join(<1. .127,211. .401,635. .>742)
                     /gene="Mn-SOD"
                     /EC_number="1.15.1.1"
                     /codon_start=1
                     /product="manganese-superoxide dismutase"
                     /protein_id="AAB16792.1"
                     /db_xref="PID:g1519024"
                     /db_xref="GI:1519024"
                     /translation="LHHKKHHQTYVNALNAAEQAYAKATTPKERIALQSALKFNGGGH
                     INHSLFWKNLAPAKSEGKGNGGAIRDGPLKSAIEQNFGSVEAFIKEFNATTASIQGSG
                     WGWLGLNPATKRLEITTTANQDPLLTHVPIIGVDIWEHAF"
BASE COUNT      142 a    262 c    203 g    135 t
ORIGIN


alignment_scores:
            Quality:   80.00              Length:    224
              Ratio:  0.690                Gaps:      9
 Percent Similarity:  51.786   Percent Identity: 17.411

alignment_block:
 US-09-068-804-7 x GTU56117   ..

 Align seg 1/1  to: GTU56117  from: 1  to: 742

       9 AlaSerProHisProGlyIleValAlaGluArgProGlnThrProSerAl 25
         ::::::|||:::|||         :::   |||:::   :::|||||::
      59 CCTACGCCAAGGCCACCA......CCCCCAAGGAGCGCATCGCCCTCCAG 102

      25 aSerGluHisValGluThrAlaValValProSerThrThrGluHisArgG 42
         ::::             ||||||||||||:::   :::   :::   |||
     103 TCCGCCCTCAAGTTCAACGGCGGTGGTGAGTGAATCTCGCCAAAGTCGGC 152

      42 lyThrAspIle.....................IleSerLeuSerGlnAla 51
           |||:::::::                    :::   |||   ::::::
     153 TCACCGCGACGCGATTACGAATCCACTGACGCCTTCGGCTGCGGCGGGGG 202

      52 AlaThrLysIleHisGlnAlaGlnGlnThrLeuGlnSerThrProPro.. 67
         ::::::   ||||||:::   :::   |||:::   :::   |||
     203 CGCGCCAGGTCACATCAACCACTCCCTCTTCTGGAAAAACCTCGCCCCCG 252

      68 ....IleSerGluGluAsnAsnAspGluArgThrLeuAlaArgGlnGlnL 83
             :::      :::::::::      |||      :::|||::::::
     253 CCAAGTCCGAGGGCAAGGGAAACGGCGGCGCCATCAGGGACGGCCCCCTC 302

      83 euThrSerSerLeuAsnAlaLeu...AlaLysSerGlyValSerLeuSer 98
           :::::::::   ::::::|||      :::   |||:::      :::
     303 AAGTCGGCGATCGAGCAGAACTTTGGTTCCGTCGAGGCCTTCATCAAGGA 352

      99 AlaGluGlnAsnGluAsnLeuArgSerAla..................... 108
         ::::::::::::::   ::::::::::::
     353 GTTCAACGCGACCACCGCGTCCATCCAGGGCTCCGGCTGGGGCTGGCTCG 402

     109 .............PheSerAlaProThrSerAlaLeuPheSerAlaSerP 121
                     ::::::   |||:::|||   ||||||:::   :::
     403 TGAGTTGCTACAGTACCCCTTACCGTCCTCTCTTCTCTTCTGC...GATC 449

     121 roMetAlaGlnPro..................................... 125
```

```
         :::|||:::|||
     450 GGAAGGCGCATCCGGTGTTATGGTCCGCGCGCTGCACGCGTGTCTCGCTT 499

     126 .............ArgThrThrIleSerAspAlaGluIleTrpAspMetVa 138
                     |||:::::::   :::|||::::::::::   ::::::::
     500 CGGCTCCGTTGTCGAAGCGCGGATACCGATACCGATCTGCTGCAAATCAT 549

     138 lSerGlnAsnIleSerAlaIleGlyAspSerTyrLeuGlyValTyrGluA 155
         ::::   :::|||          |||
     550 TTGCGCTGCGATAGAGCTCCGGGGG......................... 574

     155 snValValAlaValTyrThrAspPheTyrGlnAlaPheSerAspIleLeu 171
                             :::   ::::::|||      |||
     575 ......................CTTTTACCCTCGTTCCATGTGATAAGA 601

     172 SerLysMetGlyGlyTrp......LeuLeuProGlyLysAspGlyAsnTh 186
         :::::::::   ||||||      |||   :::|||   :::   :::||
     602 CCGCGTCTGACCGGGTGGTGTTTTCTATCGCAGGGTCTCAACCCGGCGAC 651

     186 rValLysLeuAspValThrSer 193
         |   :::|||::::::|||:::
     652 CAAGCGCCTCGAGATCACGACG 673
```

```
seq_name: gb_bal:BJA225023

seq_documentation_block:
LOCUS       BJA225023    5865 bp    DNA             BCT       13-MAR-1998
DEFINITION  Bradyrhizobium japonicum ragC and ragD genes.
ACCESSION   AJ225023
NID         g2961292
VERSION     AJ225023.1  GI:2961292
KEYWORDS    ragC gene; ragD gene.
SOURCE      Bradyrhizobium japonicum.
  ORGANISM  Bradyrhizobium japonicum
            Bacteria; Proteobacteria; alpha subdivision; Bradyrhizobium group;
            Bradyrhizobium.
REFERENCE   1  (bases 1 to 5865)
  AUTHORS   Narberhaus,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-1998) Narberhaus F., ETH-Zentrum,
            Mikrobiologisches Institut, Schmelzbergstrasse 7, CH-8092 Zuerich,
            SWITZERLAND
REFERENCE   2  (bases 1 to 5865)
  AUTHORS   Narberhaus,F., Krummenacher,P., Fischer,H.M. and Hennecke,H.
  TITLE     Three disparately regulated genes for sigma 32-like transcription
            factors in Bradyrhizobium japonicum
  JOURNAL   Mol. Microbiol. 24 (1), 93-104 (1997)
  MEDLINE   97285751
FEATURES             Location/Qualifiers
     source          1. .5865
                     /organism="Bradyrhizobium japonicum"
                     /db_xref="taxon:375"
     gene            61. .3243
                     /gene="ragC"
     CDS             61. .3243
                     /gene="ragC"
                     /function="cation efflux or multidrug resistance protein"
                     /codon_start=1
                     /transl_table=11
                     /product="RagC"
                     /protein_id="CAA12353.1"
                     /db_xref="PID:e1270775"
                     /db_xref="PID:g2961293"
                     /db_xref="GI:2961293"
                     /db_xref="SPTREMBL:O69238"
                     /translation="MGIVRFALRLPHTFYVLAALILFLGGIAIRSMPTDIFPEIRIPV
                     VTVIWRYTGLSTPEMEQRVSTYSQYSISANVTGIKNIEAQTLNGLSVQKIYFQPDVNL
                     DLAISQIVSATNAIRALMPPGIQPPIIVQFNASSVPVLQLSLESNSLNEQQLYDFGIY
                     RVRQQLAPVPGVTLPTPAGGKYRQIMVDIDPNKLLSRGLTPLDIVNAVNTQNLTLPTG
                     TTKIGDTQYTVRTNATPATIQDLNMIPVKFANGATIFLKDVAQVRDGAQVQQNIVRED
                     GHRAVLLSVIKNGNASTLAVVNGVKAALTSIRASAPAGLKINELFDQSIFVTHSVNGV
                     LREGAIAAGLTALMILVFLGSWRSTLVVLISIPLAMLSSLVVLYFLGETLNTMTLGGL
                     ALAVGILVDDSTVTIENTPALDRRGHAVVGSDPARSRRDRYSDVGLHARDQLRVHLGR
```

```
                FKEHRKAKDSALQSIESKMLELTNRLQESQEEIQIMIKEKEEMKRVQEALQIERDQLK
                ENTKEIVAKMKESQEKEYQFLKMTAVNETQEKMCEIEHLKEQFETQKLNLENIETENI
                RLTQILHENLEEMRSVTKERDDLRSVEETLKVERDQLKENLRETITRDLEKQEELKIV
                HMHLKEHQETIDKLRGIVSEKTNEISNMQKDLEHSNDALKAQDLKIQEELRIAHMHLK
                EQQETIDKLRGIVSEKTDKLSNMQKDLENSNAKLQEKIQELKANEHQLITLKKDVNET
                QKKVSEMEQLKKQIKDQSLTLSKLEIENLNLAQELHENLEEMKSVMKERDNLRRVEET
                LKLERDQLKESLQETKARDLEIQQELKTARMLSKEHKETVDKLREKISEKTIQISDIQ
                KDLDKSKDELQKKIQELQKKELQLLRVKEDVNMSHKKINEMEQLKKQFEPNYLCKCEM
                DNFQLTKKLHESLEEIRIVAKERDELRRIKESLKMERDQFIATLREMIARDRQNHQVK
                PEKRLLSDGQQHLMESLREKCSRIKELLKRYSEMDDHYECLNRLSLDLEKEIEFHRIM
                KKLKYVLSYVTKIKEEQHECINKFEMDFIDEVEKQKELLIKIQHLQQDCDVPSRELRD
                LKLNQNMDLHIEEILKDFSESEFPSIKTEFQQVLSNRKEMTQFLEEWLNTRFDIEKLK
                NGIQKENDRICQVNNFFNNRIIAIMNESTEFEERSATISKEWEQDLKSLKEKNEKLFK
                NYQTLKTSLASGAQVNPTTQDNKNPHVTSRATQLTTEKIRELENSLHEAKESAMHKES
                KIIKMQKELEVTNDIIAKLQAKVHESNKCLEKTKETIQVLQDKVALGAKPYKEEIEDL
                KMKLVKIDLEKMKNAKEFEKEISATKATVEYQKEVIRLLRENLRRSQQAQDTSVISEH
                TDPQPSNKPLTCGGGSGIVQNTKALILKSEHIRLEKEISKLKQQNEQLIKQKNELLSN
                NQHLSNEVKTWKERTLKREAHKQVTCENSPKSPKVTGTASKKKQITPSQCKERNLQDP
                VPKESPKSCFFDSRSKSLPSPHPVRYFDNSSLGLCPEVQNAGAESVDSQPGPWHASSG
                KDVPECKTQ"
BASE COUNT     3429 a   1263 c   1677 g   1888 t
ORIGIN


alignment_scores:
            Quality:   92.50            Length:    388
              Ratio:  0.503              Gaps:     16
 Percent Similarity:  47.423   Percent Identity: 18.814

alignment_block:
 US-09-068-804-7 x HSCENPE   ..

 Align seg 1/1  to: HSCENPE  from: 1  to: 8257

     27 GluHisValGluThrAlaValValProSerThrThrGluHisArgGlyTh 43
        |||:::::::|||:::         \\\      \\\   :::|||   ::
   2770 GAACAATTAGAAAATAGAGATTCTCCGCTGCAAACTGTAGAAAGGGAGAA 2819

     43 rAspIleIleSerLeuSerGlnAlaAlaThrLysIleHisGlnAlaGlnG 60
        :   :::|||:::                          ::::::   ||||
   2820 AACACTGATTACT.......................GAGAAACTGCAGC 2845

     60 lnThrLeuGlnSerThrProProIleSerGluGluAsnAsnAspGluArg 76
        ||||||||:::   :::   ::::::::::::|||:::::|||   :::
   2846 AAACTTTAGAAGAAGTAAAAACTTTAACTCAAGAAAAAGATGATCTAAAA 2895

     77 ThrLeu..................AlaArgGlnGlnLeuThrSerSerLe 87
           |||                   :::|||:::||||||:::|||:::::
   2896 CAACTCCAAGAAAGCTTGCAAATTGAGAGGGACCAACTCAAAAGTGATAT 2945

     87 uAsnAlaLeuAlaLysSerGlyValSerLeuSerAlaGluGlnAsnGluA 104
        ::::           :::   |||:::::::::   :::   :::||||:
   2946 TCAC............GATACTGTTAACATGAATATAGATACTCAAGAAC 2983

    104 snLeuArgSerAlaPheSerAlaProThrSerAlaLeuPheSerAlaSer 120
        ::||||||:::|||:::
   2984 AATTACGAAATGCTCTT.............................GAG 3003

    121 ProMetAlaGlnProArgThrThrIleSer......AspAlaGluIleTr 135
        ::::::   |||   :::   ||||||:::         :::   |||
   3004 TCTCTGAAACAACATCAAGAAACAATTAATACACTAAAATCGAAAATTTC 3053

    135 pAspMetValSerGlnAsnIleSerAlaIleGlyAspSerTyrLeuGlyV 152
         :::   ||||||:::|||                        |||   :
   3054 TGAGGAAGTTTCCAGGAAT.....................TTGCATA 3079

    152 alTyrGluAsnValValAlaValTyrThrAspPheTyrGlnAlaPheSer 168
        ::   ||||||:::   ::::::     :::|||   |||
   3080 TGGAGGAAAATACAGGAGAAACTAAAGATGAATTTCAGCAAAAGATGGTT 3129

    169 AspIleLeuSerLysMetGlyGlyTrpLeuLeuProGlyLysAspGlyAs 185
        :::|||      |||   :::         |||   :::|||:::   ::
   3130 GGCATAGATAAAAAACAGGAT.........TTGGAAGCTAAAAATACCCA 3170

    185 nThrValLysLeuAspValThrSer..................... 193
        :|||::::::   ||||||::::::
   3171 AACACTAACTGCAGATGTTAAGGATAATGAGATAATTGAGCAACAAAGGA 3220

    194 ..........................LeuLysAsnAspLeuAsn 199
                                   |||::::::   |||:::
   3221 AGATATTTTCTTTAATACAGGAGAAAAATGAACTCCAACAAATGTTAGAG 3270

    200 SerLeuValAsnLysTyrAsnGlnIleAsnSerAsnThrValLeuPhePr 216
        |||:::::::::      :::|||:::::::::::
   3271 AGTGTTATAGCAGAAAAGGAACAATTGAAGACTGAC.............. 3306

    216 oAlaGlnSerGlySerGlyValLysValAlaThrGluAlaGluAlaArgG 233
           :::   :::      :::::::::::::::::::::|||   |||
   3307 .CTAAAGGAAAATATTGAAATGACCATTGAAAACCAGGAAGAATTAAGAC 3355

    233 lnTrpLeuSerGluLeu.............................. 238
           :::|||||
   3356 TTCTTGGGGATGAACTTAAAAAGCAACAAGAGATAGTTGCACAAGAAAAG 3405

    239 .......................AsnLeuProAsnSerCys........ 244
                                :::|||:::   :::|||
   3406 AACCATGCCATAAAGAAAGAAGGAGAGCTTTCTAGGACCTGTGACAGACT 3455

    244 ............................................... 244

   3456 GGCAGAAGTTGAAGAAAAACTAAAGGAAAAGAGCCAGCAACTCCAAGAAA 3505

    245 ..............LeuLysSerTyrGlySerGlyTyrValValThrVal 256
                     |||:::                      |||
   3506 AACAGCAACAACTTCTTAAT.....................GTACAAGAA 3534

    257 AspLeuThrProLeuGlnLysMetValGlnAspIleAspGlyLeuGlyAl 273
        :::::::::   :::||||||:::::::::::::|||:::::|||   ::
   3535 GAGATGAGTGAGATGCAGAAAAAGATTAATGAAATAGAGAATTTAAAGAA 3584

    273 aProGlyLysAspSerLysLeuGluMetAspAsnAlaLysTyrGlnAlaT 290
        :      |||:::      |||   ::::::
   3585 TGAATTAAAGAACAAAGAATTGACATTGGAA.................... 3615

    290 rpGlnSerGlyPheLysAlaGlnGluGluAsnMetLysThrThrLeu... 305
                              :::   |||:::   :::      |||
   3616 ...................CATATGGAAACAGAGAGGCTTGAGTTGGCT 3645

    306 GlnThrLeuThrGlnLysTyrSerAsnAlaAsnSerLeuTyrAspAsnLe 322
        |||:::|||:::::::::|||   ::::::::::|||:::      :::
   3646 CAGAAACTTAATGAAAATTATGAGGAAGTGAAATCTATAACCAAAGAA.. 3693

    322 uValLysValLeuSerSerThrIleSerSerSerLeuGluThrAlaLysS 339
           |||||||||         :::      |||:::||||||:::::::
   3694 .AGAAAAGTTCTA...AAGGAATTACAGAAGTCATTTGAAACAGAGAGAG 3739

    339 erPheLeuGlnGly 343
        ::   |||:::|||
   3740 ACCACCTTAGAGGA 3753

seq_name: gb_pl1:GTU56117

seq_documentation_block:
LOCUS       GTU56117      742 bp    DNA             PLN       08-OCT-1996
DEFINITION  Ganoderma tsugae manganese-superoxide dismutase (Mn-SOD) gene,
            partial cds.
ACCESSION   U56117
NID         g1519023
VERSION     U56117.1  GI:1519023
KEYWORDS    .
SOURCE      Ganoderma tsugae strain=RSH J2.
  ORGANISM  Ganoderma tsugae
            Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Aphyllophorales;
            Ganodermataceae; Ganoderma.
REFERENCE   1  (bases 1 to 742)
  AUTHORS   Wang,H.-F.
```

```
Align seq 1/1  to: DROPUM  from: 1  to: 6833

    11 ProHisProGlyIleValAlaGluArgProGlnThrProSer......Al 25
       |||::::::|||:::   :::   ||||||   |||||||||      ::
  3417 CCGCAACAGGGAACGCAGCCGCGCCGCCCCCTCACCCCCTCGCAGCAGGG 3466

    25 aSerGluHisValGluThrAlaValValProSerThrThrGluHisArgG 42
       ::::|||:::          |||:::|||:::      :::|||   |
  3467 TGCCGAGAATCAGCCGTATCAGGTCATCCCGGCATTCCTCGATCACACGG 3516

    42 lyThrAspIleIleSerLeuSerGlnAlaAlaThrLysIleHis..... 56
       ||:::   :::::::::   :::::::::::|||   ::::::
  3517 GCTCCTTGCTGATGGGAGGACCTCGCACCGGGACGCCGATGCGTCTGGTT 3566

    56 ..................................................  56

  3567 AGCCCCGCCCCCGTTCTGGTGCCCCCGGGCGCTACCCGTGCCGGCCCCCC 3616

    57 .......GlnAlaGlnGlnThrLeuGlnSerThrProProIleSerGluG 71
              |||::::::|||      |||:::   |||:::::::::::::
  3617 GCCCCCGCAGGGGCCACAGCTGTATCAGCCGCAGCCGCAGACGGCCCAAC 3666

    71 luAsnAsnAspGluArgThrLeuAlaArgGlnGlnLeuThrSerSerLeu 87
       ::|||              |||      ||||||      ||||||:::
  3667 AGAAT..............CTCTACTCGCAGCAGAATGGATCCAGTGTC 3701

    88 AsnAlaLeuAlaLysSerGlyValSerLeuSerAlaGluGlnAsnGluAs 104
       ::::::|||||   :::      ||||||::::::
  3702 GGAGGCCTCGCCTTGAACACGAGCTCGTTGACGGGTCGC........... 3740

   104 nLeuArgSerAlaPheSerAlaProThrSerAlaLeuPheSerAlaSerP 121
          |||::::::|||:::   :::|||||||||   ||||||:::|||:
  3741 ....CGCGACTCCTTCGACCGCAGCACCTCCGCC...TTCAGTCCCTCGA 3783

   121 roMetAlaGlnProArgThrThrIleSerAspAla......GluIleTrp 135
       ::|||:::   :::   :::   ::::::::::|||      ::::::
  3784 CCATGGACTACACCAGCAGCGGTGTGGCAGCGGCCGCCAATGCGGTGAAC 3833

   136 AspMetValSerGlnAsnIleSerAlaIleGlyAspSerTyrLeuGlyVa 152
       :::   |||:::|||:::   :::|||   :::
  3834 AGCACAGTGGCCCAGGCAGCAGCAGCTGCCGCAGC............... 3868

   152 lTyrGluAsnValValAlaValTyrThrAspPheTyrGlnAlaPheSerA 169

  3868 ..................................................  3868

   169 spIleLeuSerLysMetGlyGlyTrpLeuLeuProGlyLysAspGlyAsn 185
              |||:::   :::::::|||   ::::::|||   |||   :::
  3869 ........AGCCGCCGCAGCGCGTGGCAAGTGGCCGGGAGCGATGTCGGG 3910

   186 ThrValLysLeuAspValThrSerLeuLysAsnAsp......LeuAsnSe 200
       :::   :::   ::::::   |||      ::::::      ::::::
  3911 AGCGGCCAGTGGAGCCTACGGAGCCCTGGGAGCGGGCAATGCTTCGGCCA 3960

   200 rLeuValAsnLysTyrAsnGlnIleAsnSer...AsnThrValLeuPheP 216
              ::::::   |||:::   ::::::   ::::::::|||   |
  3961 GTCCCCTGGGCGCACCAATCACGCCGCCGCCATCGGCGCAATCCTGTCTC 4010

   216 roAlaGlnSerGlySerGlyValLysVal..............AlaThr 227
       ||:::|||||||||:::      :::|||              ||||||
  4011 CTGGGCAGTCGGGCACCTGGAGCCGAGTCCCGCCAGCGGCAGCAGCAACA 4060

   228 GluAlaGluAlaArgGlnTrpLeuSerGluLeuAsnLeuProAsnSerCy 244
          |||::::::|||:::|||                 ::::::::||
  4061 ACAGCAGCTGGCCGCCGTTGG..............TCTGCCGGCGACTG 4095

   244 sLeuLysSerTyrGlySerGly.................TyrValValT 255
       |      :::   ||||||||                  ||||||
  4096 CAGCAGCTGCTCAGGCAGCGGTGGCCGCGGCTGCCAACAATATGTT.... 4141

   255 hrValAspLeuThrProLeuGlnLysMetValGlnAspIleAspGlyLeu 271
                                 :::|||::::::|||   |||
  4142 .......................CGGATCCAACAGCTCGATCTTCTC 4165

   272 GlyAlaProGlyLysAspSerLysLeuGluMetAspAsnAlaLysTyrGl 288
          :::|||||
  4166 GAATCCCCTGGC..................................... 4177

   288 nAlaTrpGlnSerGlyPheLysAla 296
             :::||||||:::::::::
  4178 .......CATTCCGGGTACCGCAGC 4195
```

```
seq_name: gb_pr2:HSCENPE

seq_documentation_block:
LOCUS       HSCENPE      8257 bp    mRNA            PRI       10-JAN-1993
DEFINITION  H.sapiens CENP-E mRNA.
ACCESSION   Z15005
NID         g29864
VERSION     Z15005.1  GI:29864
KEYWORDS    kinetochore motor.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 8257)
  AUTHORS   Yen,T.J., Li,G., Schaar,B.T., Szilak,I. and Cleveland,D.W.
  TITLE     CENP-E is a putative kinetochore motor that accumulates just before
            mitosis
  JOURNAL   Nature 359 (6395), 536-539 (1992)
  MEDLINE   93024922
REFERENCE   2  (bases 1 to 8257)
  AUTHORS   Cleveland,D.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-AUG-1992) Don W. Cleveland, Biological Chemistry,
            Johns Hopkins University, School of Medicine, 725 N. Wolfe St.,
            Baltimore, Maryland, 21205, USA
FEATURES             Location/Qualifiers
     source          1. .8257
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CENPE"
     CDS             91. .8082
                     /codon_start=1
                     /product="CENP-E"
                     /protein_id="CAA78727.1"
                     /db_xref="PID:g29865"
                     /db_xref="GI:29865"
                     /db_xref="SWISS-PROT:Q02224"
                     /translation="MAEEGAVAVCVRVRPLNSREESLGETAQVYWKTDNNVIYQVDGS
                     KSFNFDRVFHGNETTKNVYEEIAAPIIDSAIQGYNGTIFAYGQTASGKTYTMMGSEDH
                     LGVIPRAIHDIFQKIKKFPDREFLLRVSYMEIYNETITDLLCGTQKMKPLIIREDVNR
                     NVYVADLTEEVVYTSEMALKWITKGEKSRHYGETKMNQRSSRSHTIFRMILESREKGE
                     PSNCEGSVKVSHLNLVDLAGSERAAQTGAAGVRLKEGCNINRSLFILGQVIKKLSDGQ
                     VGGFINYRDSKLTRILQNSLGGNPKTRIICTITPVSFDETLTALQFASTAKYMKNTPY
                     VNEVSTDEALLKRYRKEIMDLKKQLEEVSLETRAQAMEKDQLAQLLEEKDLLQKVQNE
                     KIENLTRMLVTSSSLTLQQELKAKRKRRVTWCLGKINKMKNSNYADQFNIPTNITTKT
                     HKLSINLLREIDESVCSESDVFSNTLDTLSEIEWNPATKLLNQENIESELNSLRADYD
                     NLVLDYEQLRTEKEEMELKLKEKNDLDEFEALERKTKKDQEMQLIHEISNLKNLVKHR
                     EVYNQDLENELSSKVELLREKEDQIKKLQEYIDSQKLENIKMDLSYSLESIEDPKQMK
                     QTLFDAETVALDAKRESAFLRSENLELKEKMKELATTYKQMENDIQLYQSQLEAKKKM
                     QVDLEKELQSAFNEITKLTSLIDGKVPKDLLCNLELEGKITDLQKELNKEVEENEALR
                     EEVILLSELKSLPSEVERLRKEIQDKSEELHIITSEKDKLFSEVVHKESRVQGLLEEI
                     GKTKDDLATTQSNYKSTDQEFQNFKTLHMDFEQKYKMVLEENERMNQEIVNLSKEAQK
                     FDSSLGALKTELSYKTQELQEKTREVQERLNEMEQLKEQLENRDSPLQTVEREKTLIT
                     EKLQQTLEEVKTLTQEKDDLKQLQESLQIERDQLKSDIHDTVNMNIDTQEQLRNALES
                     LKQHQETINTLKSKISEEVSRNLHMEENTGETKDEFQQKMVGIDKKQDLEAKNTQTLT
                     ADVKDNEIIEQQRKIFSLIQEKNELQQMLESVIAEKEQLKTDLKENIEMTIENQEELR
                     LLGDELKKQQEIVAQEKNHAIKKEGELSRTCDRLAEVEEKLKEKSQQLQEKQQQLLNV
                     QEEMSEMQKKINEIENLKNELKNKELTLEHMETERLELAQKLNENYEEVKSITKERKV
                     LKELQKSFETERDHLRGYIREIEATGLQTKEELKIAHIHLKEHQETIDELRRSVSEKT
                     AQIINTQDLEKSHTKLQEEIPVLHEEQELLPNVKKVSETQETMNELELLTEQSTTKDS
                     TTLARIEMERLRLNEKFQESQEEIKSLTKERDNLKTIKEALEVKHDQLKEHIRETLAK
                     IQESQSKQEQSLNMKEKDNETTKIVSEMEQFKPKDSALLRIEIEMLGLSKRLQESHDE
                     MKSVAKEKDDLQRLQEVLQSESDQLKENIKEIVAKHLETEEELKVAHCCLKEQEETIN
                     ELRVNLSEKETEISTIQKQLEAINDKLQNKIQEIYEKEEQLNIKQISEVQENVNELKQ
```

```
                RRPLTPSQQGAENQPYQVIPAFLDHTGSLLMGGPRTGTPMRLVSPAPVLVPPGATRAG
                PPPPQGPQLYQPQPQTAQQNLYSQQNGSSVGGLALNTSSLTGRRDSFDRSTSAFSPST
                MDYTSSGVAAAANAVNSTVAQAAAAAAAAAAARGKWPGAMSGAASGAYGALGAGNASA
                SPLGAPITPPPSAQSCLLGSRAPGAESRQRQQQQQQLAAVGLPATAAAAQAAVAAAAN
                NMFGSNSSIFSNPLAIPGTAAVAAAAAAAAAANSRQVAATAAAAAAVAAAAGGVGGAP
                QPGRSRLLEDFRNQPYPNLQLRDLANHIVEFSQDQHGSRFIQQKLERATAAEKQMVFS
                EILAAAYSLMTDVFGNYVIQKFFEFGTPEQKNTLGMQVKGHVLQLALQMYGCRVIQKA
                LESISPEQQQEIVHELDGHVLKCVKDQNGNHVVQKCIECVDPVALQFIINAFKGQVYS
                LSTHPYGCRVIQRILEHCTAEQTTPILDELHEHTEQLIQDQYGNYVIQHVLEHGKQED
                KSILINSVRGKVLVLSQHKFASNVVEKCVTHATRGERTGLIDEVCTFNDNALHVMMKD
                QYANYVVQKMIDVSEPTQLKKLMTKIRPHMAALRKYTYGKHINAKLEKYYMKITNPIT
                VGTGAGGVPAASSAAAVSSGATSASVTACTSGSSTTTTSTTNSLASPTICSVQENGSA
                MIVEPSSPDASESSSSVVSGAVNSGLGPIGPPTNGNVVL"
misc_difference 1277
                /gene="pum"
                /note="change observed in genomic DNA"
                /db_xref="FlyBase:FBgn0003165"
                /replace="g"
exon            1377. .2159
                /gene="pum"
                /db_xref="FlyBase:FBgn0003165"
                /number=3
                /evidence=experimental
exon            2160. .2401
                /gene="pum"
                /db_xref="FlyBase:FBgn0003165"
                /number=4
                /evidence=experimental
exon            2402. .2528
                /gene="pum"
                /db_xref="FlyBase:FBgn0003165"
                /number=5
                /evidence=experimental
exon            2529. .2647
                /gene="pum"
                /db_xref="FlyBase:FBgn0003165"
                /number=6
                /evidence=experimental
exon            2648. .2936
                /gene="pum"
                /db_xref="FlyBase:FBgn0003165"
                /number=7
                /evidence=experimental
exon            2937. .3023
                /gene="pum"
                /db_xref="FlyBase:FBgn0003165"
                /number=8
                /evidence=experimental
misc_difference 2975
                /gene="pum"
                /note="change observed in genomic DNA"
                /db_xref="FlyBase:FBgn0003165"
                /replace="c"
exon            3024. .3488
                /gene="pum"
                /db_xref="FlyBase:FBgn0003165"
                /number=9
                /evidence=experimental
exon            3489. .3705
                /gene="pum"
                /db_xref="FlyBase:FBgn0003165"
                /number=10
                /evidence=experimental
exon            3706. .4983
                /gene="pum"
                /db_xref="FlyBase:FBgn0003165"
                /number=11
                /evidence=experimental
misc_difference 4351
                /gene="pum"
                /note="change observed in genomic DNA"
                /db_xref="FlyBase:FBgn0003165"
                /replace="g"
exon            4984. .5152
                /gene="pum"
                /db_xref="FlyBase:FBgn0003165"
                /number=12
                /evidence=experimental
exon            5153. .6833
                /gene="pum"
                /db_xref="FlyBase:FBgn0003165"
                /number=13
                /evidence=experimental
misc_difference 5396
                /gene="pum"
                /note="change observed in genomic DNA"
                /db_xref="FlyBase:FBgn0003165"
                /replace="g"
misc_difference 5399
                /gene="pum"
                /note="change observed in genomic DNA"
                /db_xref="FlyBase:FBgn0003165"
                /replace="a"
misc_difference 5529
                /gene="pum"
                /note="change observed in genomic DNA"
                /db_xref="FlyBase:FBgn0003165"
                /replace="g"
misc_difference 5598
                /gene="pum"
                /note="change observed in genomic DNA"
                /db_xref="FlyBase:FBgn0003165"
                /replace="a"
misc_difference 5851. .5852
                /gene="pum"
                /note="change (insertion) observed in genomic DNA"
                /db_xref="FlyBase:FBgn0003165"
                /replace="atag"
misc_difference 5878
                /gene="pum"
                /note="change observed in genomic DNA"
                /db_xref="FlyBase:FBgn0003165"
                /replace="a"
misc_difference 5986
                /gene="pum"
                /note="change observed in genomic DNA"
                /db_xref="FlyBase:FBgn0003165"
                /replace="c"
misc_difference 6053
                /gene="pum"
                /note="change observed in genomic DNA"
                /db_xref="FlyBase:FBgn0003165"
                /replace="a"
misc_difference 6058
                /gene="pum"
                /note="change observed in genomic DNA"
                /db_xref="FlyBase:FBgn0003165"
                /replace="g"
misc_difference 6175
                /gene="pum"
                /note="change observed in genomic DNA"
                /db_xref="FlyBase:FBgn0003165"
                /replace="t"
misc_difference 6466
                /gene="pum"
                /note="change observed in genomic DNA"
                /db_xref="FlyBase:FBgn0003165"

alignment_scores:
            Quality:   92.00              Length:    325
              Ratio:  0.561                Gaps:     14
 Percent Similarity:  50.462   Percent Identity: 21.538

alignment_block:
 US-09-068-804-7 x DROPUM    ..
```

```
  11 ProHisProGlyIleValAlaGluArgProGlnThrProSer......Al 25
     |||:::::::|||:::   :::   ||||||   |||||||||      ::
2698 CCGCAACAGGGAACGCAGCCGCGCCGCCCCCTCACCCCCTCGCAGCAGGG 2747

  25 aSerGluHisValGluThrAlaValValProSerThrThrGluHisArgG 42
     ::::|||:::           |||:::|||:::      :::|||   |
2748 TGCCGAGAATCAGCCGTATCAGGTCATCCCGGCATTCCTCGATCACACGG 2797

  42 lyThrAspIleIleSerLeuSerGlnAlaAlaThrLysIleHis..... 56
     ||:::   :::::::::   ::::::::::::|||   ::::::
2798 GCTCCTTGCTGATGGGAGGACCTCGCACCGGGACGCCGATGCGTCTGGTT 2847

  56 ..................................................... 56

2848 AGCCCCGCCCCCGTTCTGGTGCCCCCGGGCGCTACCCGTGCCGGCCCCCC 2897

  57 .......GlnAlaGlnGlnThrLeuGlnSerThrProProIleSerGluG 71
           |||::::::|||      |||:::   |||:::::::::::::
2898 GCCCCCGCAGGGCCCACAGCTGTATCAGCCGCAGCCGCAGACGGCCCAAC 2947

  71 luAsnAsnAspGluArgThrLeuAlaArgGlnGlnLeuThrSerSerLeu 87
     ::|||              |||      ||||||      ||||||:::
2948 AGAAT..............CTCTACTCGCAGCAGAATGGATCCAGTGTC 2982

  88 AsnAlaLeuAlaLysSerGlyValSerLeuSerAlaGluGlnAsnGluAs 104
     ::::::||||||   :::      ||||||::::::
2983 GGAGGCCTCGCCTTGAACACGAGCTCGTTGACGGGTCGC........... 3021

 104 nLeuArgSerAlaPheSerAlaProThrSerAlaLeuPheSerAlaSerP 121
         |||::::::|||:::   :::|||||||||   ||||||:::|||:
3022 ....CGCGACTCCTTCGACCGCAGCACCTCCGCC...TTCAGTCCCTCGA 3064

 121 roMetAlaGlnProArgThrThrIleSerAspAla......GluIleTrp 135
     ::|||:::   :::   :::   ::::::::::|||      ::::::
3065 CCATGGACTACACCAGCAGCGGTGTGGCAGCGGCCGCCAATGCGGTGAAC 3114

 136 AspMetValSerGlnAsnIleSerAlaIleGlyAspSerTyrLeuGlyVa 152
     :::   |||:::|||:::   :::|||   :::
3115 AGCACAGTGGCCCAGGCAGCAGCAGCTGCCGCAGC.............. 3149

 152 lTyrGluAsnValValAlaValTyrThrAspPheTyrGlnAlaPheSerA 169

3149 ..................................................... 3149

 169 spIleLeuSerLysMetGlyGlyTrpLeuLeuProGlyLysAspGlyAsn 185
             |||:::   :::::::|||   ::::::|||   |||   :::
3150 ........AGCCGCCGCAGCGCGTGGCAAGTGGCCGGGAGCGATGTCGGG 3191

 186 ThrValLysLeuAspValThrSerLeuLysAsnAsp......LeuAsnSe 200
     :::   :::   ::::::   |||      ::::::      ::::::
3192 AGCGGCCAGTGGAGCCTACGGAGCCCTGGGAGCGGGCAATGCTTCGGCCA 3241

 200 rLeuValAsnLysTyrAsnGlnIleAsnSer...AsnThrValLeuPheP 216
             ::::::   |||:::   ::::::   ::::::::::|||   |
3242 GTCCCCTGGGCGCACCAATCACGCCGCCGCCATCGGCGCAATCCTGTCTC 3291

 216 roAlaGlnSerGlySerGlyValLysVal..............AlaThr 227
     ||:::|||||||||:::       :::|||                ||||||
3292 CTGGGCAGTCGGGCACCTGGAGCCGAGTCCCGCCAGCGGCAGCAGCAACA 3341

 228 GluAlaGluAlaArgGlnTrpLeuSerGluLeuAsnLeuProAsnSerCy 244
        |||:::::|||:::|||                 ::::::::||
3342 ACAGCAGCTGGCCGCCGTTGG..............TCTGCCGGCGACTG 3376

 244 sLeuLysSerTyrGlySerGly..................TyrValValT 255
     |       :::   |||||||||                  ||||||
3377 CAGCAGCTGCTCAGGCAGCGGTGGCCGCGGCTGCCAACAATATGTT.... 3422

 255 hrValAspLeuThrProLeuGlnLysMetValGlnAspIleAspGlyLeu 271
                                  :::|||:::::::|||   |||
3423 ........................CGGATCCAACAGCTCGATCTTCTC 3446

 272 GlyAlaProGlyLysAspSerLysLeuGluMetAspAsnAlaLysTyrGl 288
```

```
         :::||||||
3447 GAATCCCCTGGC..................................... 3458

 288 nAlaTrpGlnSerGlyPheLysAla 296
           :::|||||||:::::::::
3459 .......CATTCCGGGTACCGCAGC 3476
```

```
seq_name: gb_in1:DROPUM

seq_documentation_block:
LOCUS       DROPUM       6833 bp    mRNA            INV       18-SEP-1997
DEFINITION  Drosophila melanogaster pumilio protein mRNA, complete cds.
ACCESSION   L07943
NID         g158190
VERSION     L07943.1  GI:158190
KEYWORDS    pumilio protein.
SOURCE      Drosophila melanogaster 0 to 8 hours embryo cDNA to mRNA.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 6833)
  AUTHORS   Barker,D.D., Wang,C., Moore,J., Dickinson,L.K. and Lehmann,R.
  TITLE     Pumilio is essential for function but not for distribution of the
            Drosophila abdominal determinant Nanos
  JOURNAL   Genes Dev. 6, 2312-2326 (1992)
  MEDLINE   93093466
FEATURES             Location/Qualifiers
     source          1. .6833
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /dev_stage="0 to 8 hours"
                     /tissue_type="embryo"
                     /map="85CD 3-48.5"
     gene            1. .6833
                     /gene="pum"
                     /note="pumilio"
                     /allele=""
                     /db_xref="FlyBase:FBgn0003165"
     exon            1. .240
                     /gene="pum"
                     /db_xref="FlyBase:FBgn0003165"
                     /number=1
                     /evidence=experimental
     exon            241. .1376
                     /gene="pum"
                     /db_xref="FlyBase:FBgn0003165"
                     /number=2
                     /evidence=experimental
     CDS             1044. .5645
                     /gene="pum"
                     /note="maternally required for embryonic abdomen
                     formation"
                     /codon_start=1
                     /evidence=experimental
                     /product="pumilio protein"
                     /protein_id="AAB59189.1"
                     /db_xref="PID:g158191"
                     /db_xref="GI:158191"
                     /translation="MKFLGGNDDRNGRGGVGVGTDAIVGSRGGVSQDAADAAGAAAAA
                     AVGYVFQQRPSPGGVGVGVGGVGGGVPGVGAVGSTLHEAAAAEYAAHFAQKQQQTRWA
                     CGDDGHGIDNPDKWKYNPPMNPANAAPGGPPGNGSNGGPGAIGTIGMGSGLGGGGGGG
                     AGGGNNGGSGTNGGLHHQSMAAAAANMAAMQQAAALAKHNHMISQAAAAVAAQQQHQH
                     PHQQHPQQQQQQQQQAQNQGHPHHLMGGGNGLGNGNGLGIQHPGQQQQQQQQQQQQQHP
                     GQYNANLLNHAAALGHMSSYAQSGGSMYDHHGGAMHPGMNGGMPKQQPLGPPGAGGPQ
                     DYVYMGGQTTVPMGAAMMPPQNQYMNSSAVAAANRNAAITTSTAKKLWEKSDGKGVSS
                     STPGGPLHPLQIPGIGDPSSVWKDHTWSTQGENILVPPPSRAYAHGGASDTSNSGNAG
                     ILSPRDSTCAKVVEYVFSGSPTNKDSSLSGLEPHLRNLKFDDNDKSRDDKEKANSPFD
                     TNGLKKDDQVTNSNGVVNGIDDDKGFNRTPGSRQPSPAEESQPRPPNLLFPPLPFNHM
                     LMDHGQGMGGGLGGVVGSGNGVGGGSGGGGAGGAYAAHQQMAAQMSQLQPPMMNGVGG
                     GMPMAAQSPMLNHQAAGPNHMESPGNLLQQQNFDVQQLFRSQNPGLAAVATNAAAAAA
                     AAAAATSAASAAAAVGAPPVPNGSLQQSQQQQQQQQQQQQQQQMHMAAASQQFLAAQQ
                     QAQNAAYAAQQATSYVINPGQEAAPYMGMIAAAQMPYYGVAPWGMYPGNLIPQQGTQP
```

```
ACCESSION   X62589
NID         g8393
VERSION     X62589.1  GI:8393
KEYWORDS    pumilio gene.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea;
            Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 6120)
  AUTHORS   Macdonald,P.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-OCT-1991) P.M. Macdonald, Stanford Univeristy,
            Department of Biological Sciences, Stanford, CA 94305-5020, USA
REFERENCE   2  (bases 1 to 6120)
  AUTHORS   Macdonald,P.M.
  TITLE     The Drosophila pumilio gene: an unusually long transcription unit
            and an unusual protein
  JOURNAL   Development 114 (1), 221-232 (1992)
  MEDLINE   92249205
COMMENT     Maternal effect lethal defective in posterior pattern formation.
FEATURES             Location/Qualifiers
     source          1. .6120
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="III"
                     /map="85 Cytology units"
     mRNA            <1. .6120
                     /gene="pum"
                     /db_xref="FlyBase:FBgn0003165"
                     /evidence=experimental
     exon            <1. .657
                     /gene="pum"
                     /db_xref="FlyBase:FBgn0003165"
                     /number=1
                     /evidence=experimental
     gene            1. .6120
                     /gene="pum"
                     /note="pumilio"
                     /allele=""
                     /db_xref="FlyBase:FBgn0003165"
     CDS             325. .4926
                     /gene="pum"
                     /note="pumilio"
                     /codon_start=1
                     /protein_id="CAA44474.1"
                     /db_xref="PID:g8394"
                     /db_xref="GI:8394"
                     /db_xref="FLYBASE:FBgn0003165"
                     /db_xref="SWISS-PROT:P25822"
                     /translation="MKFLGGNDDRNGRGGVGVGTDAIVGSRGGVSQDAADAAGAAAAA
                     AVGYVFQQRPSPGGVGVGVGGVGGGVPGVGAVGSTLHEAAAAEYAAHFAQKQQQTRWA
                     CGDDGHGIDNPDKWKYNPPMNPANAAPGGPPGNGSNGGPGAIGTIGMGSGLGGGGGGG
                     AGGGNNGGSGTNGGLHHQSMAAAAANMAAMQQAAALAKHNHMISQAAAAVAAQQQHQH
                     PHQQHPQQQQQQQQAQNQGHPHHLMGGGNGLGNGNGLGIQHPGQQQQQQQQQQQQQHP
                     GQYNANLLNHAAALGHMSSYAQSGGSMYDHHGGAMHPGMNGGMPKQQPLGPPGAGGPQ
                     DYVYMGGQTTVPMGAAMMPPQNQYMNSAAVAAANRNAAITTSTAKKLWEKSDGKGVSS
                     STPGGPLHPLQIPGIGDPSSVWKDHTWSTQGENILVPPPSRAYAHGGASDTSNSGNAG
                     ILSPRDSTCAKVVEYVFSGSPTNKDSSLSGLEPHLRNLKFDDNDKSRDDKEKANSPFD
                     TNGLKKDDQVTNSNGVVNGIDDDKGFNRTPGSRQPSPAEESQPRPPNLLFPPLPFNHM
                     LMDHGQGMGGGLGGVVGSGNGVGGGSGGGGAGGAYAAHQQMAAQMSQLQPPMMNGVGG
                     GMPMAAQSPMLNHQAAGPNHMESPGNLLQQQNFDVQQLFRSQNPGLAAVATNAAAAAA
                     AAAAATSAASAAAAVGAPPVPNGSLQQSQQQQQQQQQQQQQQQMHMAAASQQFLAAQQ
                     QAQNAAYAAQQATSYVINPGQEAAPYMGMIAAAQMPYYGVAPWGMYPGNLIPQQGTQP
                     RRPLTPSQQGAENQPYQVIPAFLDHTGSLLMGGPRTGTPMRLVSPAPVLVPPGATRAG
                     PPPPQGPQLYQPQPQTAQQNLYSQQNGSSVGGLALNTSSLTGRRDSFDRSTSAFSPST
                     MDYTSSGVAAAANAVNSTVAQAAAAAAAAAAARGKWPGAMSGAASGAYGALGAGNASA
                     SPLGAPITPPPSAQSCLLGSRAPGAESRQRQQQQQLAAVGLPATAAAAQAAVAAAAN
                     NMFGSNSSIFSNPLAIPGTAAVAAAAAAAAAANSRQVAATAAAAAAVAAAAGGVGGAP
                     QPGRSRLLEDFRNQRYPNLQLRDLANHIVEFSQDQHGSRFIQQKLERATAAEKQMVFS
                     EILAAAYSLMTDVFGNYVIQKFFEFGTPEQKNTLGMQVKGHVLQLALQMYGCRVIQKA
                     LESISPEQQQEIVHELDGHVLKCVKDQNGNHVVQKCIECVDPVALQFIINAFKGQVYS
                     LSTHPYGCRVIQRILEHCTAEQTTPILDELHEHTEQLIQDQYGNYVIQHVLEHGKQED
                     KSILINSVRGKVLVLSQHKFASNVVEKCVTHATRGERTGLIDEVCTFNDNALHVMMKD
                     QYANYVVQKMIDVSEPTQLKKLMTKIRKNMAALRKYTYGKHINAKLEKYYMKITNPIT
                     VGTGAGGVPAASSAAAVSSGATSASVTACTSGSSTTTTSTTNSLASPTICSVQENGSA
                     MVVEPSSPDASESSSSVVSGAVNSSLGPIGPPTNGNVVL"
     exon            658. .1440
                     /gene="pum"
                     /db_xref="FlyBase:FBgn0003165"
                     /number=2
                     /evidence=experimental
     exon            1441. .1682
                     /gene="pum"
                     /db_xref="FlyBase:FBgn0003165"
                     /number=3
                     /evidence=experimental
     exon            1683. .1809
                     /gene="pum"
                     /db_xref="FlyBase:FBgn0003165"
                     /number=4
                     /evidence=experimental
     exon            1810. .1928
                     /gene="pum"
                     /db_xref="FlyBase:FBgn0003165"
                     /number=5
                     /evidence=experimental
     exon            1929. .2217
                     /gene="pum"
                     /db_xref="FlyBase:FBgn0003165"
                     /number=6
                     /evidence=experimental
     exon            2218. .2304
                     /gene="pum"
                     /db_xref="FlyBase:FBgn0003165"
                     /number=7
                     /evidence=experimental
     exon            2305. .2769
                     /gene="pum"
                     /db_xref="FlyBase:FBgn0003165"
                     /number=8
                     /evidence=experimental
     exon            2770. .2986
                     /gene="pum"
                     /db_xref="FlyBase:FBgn0003165"
                     /number=9
                     /evidence=experimental
     exon            2987. .4264
                     /gene="pum"
                     /db_xref="FlyBase:FBgn0003165"
                     /number=10
                     /evidence=experimental
     exon            4265. .4432
                     /gene="pum"
                     /db_xref="FlyBase:FBgn0003165"
                     /number=11
                     /evidence=experimental
     exon            4433. .6116
                     /gene="pum"
                     /db_xref="FlyBase:FBgn0003165"
                     /number=12
                     /evidence=experimental
BASE COUNT     1632 a   1720 c   1626 g   1142 t
ORIGIN


alignment_scores:
            Quality:   92.00              Length:    325
              Ratio:  0.561                Gaps:     14
 Percent Similarity:  50.462   Percent Identity: 21.538

alignment_block:
 US-09-068-804-7 x DMPUMI    ..

 Align seg 1/1  to: DMPUMI  from: 1  to: 6120
```

```
     948 ..................................................  948

     224 LysValAlaThrGluAlaGluAlaArgGlnTrpLeuSer..........  236
         ||||||:::   :::|||   |||:::   |||:::
     949 ...GTGGCACCGCCCGAGGAACTGCGTCGCTCACTGGCCCGCGACGCCCG  995

     237 .GluLeuAsnLeuProAsnSerCysLeuLysSerTyr.............  248
         |||::::::|||||       |||:::   :::
     996 GGAGGTATCGCTCCCGCCCCATGTACTTCGCCAGTTCCGGCACGAGTCGT 1045

     249 ........GlySerGlyTyrValValThrValAspLeuThrProLeuGln  262
                 :::|||:::            :::      |||:::
    1046 TCCGGTTGAAGATCTCGTACATC............ATGGGGCGGCTGCAC 1083

     263 LysMetValGlnAspIleAspGlyLeuGlyAlaProGlyLysAsp 277
         ::::::|||::::::|||:::         |||:::   |||
    1084 GGGCTGCTTCAGGCGCTGGATGACCCGACGCAGCCTGCGCCCGAC 1128

seq_name: gb_ba1:EFPF54

seq_documentation_block:
LOCUS       EFPF54       1900 bp    DNA             BCT       20-DEC-1989
DEFINITION  E.faecium DNA encoding P54 protein.
ACCESSION   X16421 M26048
NID         g43333
VERSION     X16421.1  GI:43333
KEYWORDS    P54 protein.
SOURCE      Enterococcus faecium.
  ORGANISM  Enterococcus faecium
            Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
            Enterococcus.
REFERENCE   1  (bases 1 to 1900)
  AUTHORS   Solioz,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-1989) M. Solioz, University of Bern, 3010 Bern,
            Switzerland
REFERENCE   2  (bases 1 to 1900)
  AUTHORS   Furst,P., Mosch,H.U. and Solioz,M.
  TITLE     A protein of unusual composition from Enterococcus faecium
  JOURNAL   Nucleic Acids Res. 17 (16), 6724 (1989)
  MEDLINE   89385998
FEATURES             Location/Qualifiers
     source          1. .1900
                     /organism="Enterococcus faecium"
                     /db_xref="taxon:1352"
     CDS             330. .1853
                     /note="P54 protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA34442.1"
                     /db_xref="PID:g43334"
                     /db_xref="GI:43334"
                     /db_xref="SWISS-PROT:P13692"
                     /translation="MLSSIALTAVGSPIAAAADDFDSQIQQQDKKIADLQNQQASAQS
                     QIEALEGQVSAINTKAQDLLTKQDTLRKESAQLKQEIKDLQERIEKREATIQKQARET
                     QVKNTSSNYIDAVLNADSLADAVGRIQAMSTIVKANQDLVQQQKEDKQAVEAKKAENE
                     AKQKELADNQAALESQKGDLLAKQADLNVLKTSLAAEQATAEDKKADLNRKKAEAEAE
                     QARIREQARLAEQARQQAAQEKAEKEAREQAAAQAAQTQALSSASTTTESSSAAQSSS
                     EESKAPESSTTEESTSTESSTTTENSSTGSSSTESSSTEESTVPESSTQESTPANTES
                     SSSSSNTNVNNNTNNSTNNSTNNSTTNNNNNNNTVTPAPTPTPTPAPAPAPNPSGSVN
                     GAAIVAEAYKYIGTPYVWGGKDPSGFDCSGFTRYVYLQVTGRDIGGWTVPQESAGTKI
                     SVSQAKAGDLLFWGSAGGTYHVAISLGGGQYIHAPQPGENVKVGSVQWYTPDFAVSM"
BASE COUNT      714 a    362 c    330 g    494 t
ORIGIN


alignment_scores:
            Quality:   89.00            Length:    261
              Ratio:  0.553              Gaps:     13
 Percent Similarity:  61.686   Percent Identity: 18.008

alignment_block:
```

```
US-09-068-804-7 x EFPF54   ..

Align seg 1/1  to: EFPF54  from: 1  to: 1900

     10 SerProHisProGlyIleValAlaGluArgProGlnThrProSerAlaSe 26
        :::|||:::::::::::::::::::|||   :::   |||   |||:::::
   1170 GCTCCTGAGTCTTCAACAACAGAAGAATCAACATCTACTGAATCTTCTAC 1219

     26 rGluHisValGluThrAlaValValProSerThrThrGluHisArgGlyT 43
        :         ::::::::::::   :::|||:::||||||      :::|
   1220 AACAACAGAAAATTCATCAACTGGCTCTTCTTCTACTGAATCATCTTCAA 1269

     43 hrAspIleIleSerLeuSerGlnAlaAlaThrLysIleHisGlnAlaGln 59
        ||:::      :::::::::::::::::::|||:::
   1270 CAGAAGAATCAACAGTTCCTGAATCTTCTACACAA.............. 1304

     60 GlnThrLeuGlnSerThrProProIleSerGluGluAsnAsnAspGluAr 76
              :::||||||||||:::   :::|||   :::::::::
   1305 .........GAGTCAACACCTGCGAATACAGAAAGCAGTTCAAGCAGCTC 1345

     76 gThrLeuAlaArgGlnGlnLeuThrSerSerLeuAsnAlaLeuAlaLysS 93
         :::   :::   :::::::   |||::::::   ::::::   :::::::
   1346 TAATACAAATGTAAATAATAATACAAACAATAGCACTAATAATAGCACTA 1395

     93 erGlyValSerLeuSerAlaGluGlnAsnGluAsnLeuArgSerAlaPhe 109
        :::::   :::   :::::::::::::|||:::|||      ::::::
   1396 ATAACAGTACAACAAATAACAACAATAACAACAATACTGTAACACCAGCT 1445

     110 SerAlaProThrSerAlaLeuPheSerAlaSerProMetAlaGlnProAr 126
         ::::::||||||::::::      ::::::::::|||   :::::::|||
   1446 CCAACACCTACACCAACGCCT...GCTCCAGCACCAGCACCAAATCCTTC 1492

     126 gThrThrIleSerAspAlaGluIleTrpAspMetValSerGlnAsnIleS 143
         :::::::::::::|||:::|||         |||:::::::::
   1493 TGGTAGTGTGAACGGCGCTGCGATT.........GTAGCAGAAGCTTACA 1533

     143 erAlaIleGlyAspSerTyrLeu.................GlyValTyr 153
         ||||||   :::|||:::                      :::   :::
   1534 AATATATTGGTACTCCATATGTATGGGGTGGTAAAGATCCAAGTGGCTTT 1583

     154 GluAsnValValAlaValTyrThrAspPheTyrGlnAlaPheSerAspIl 170
         :::         :::   :::|||   :::      ::::::   :::::
   1584 GAC......TGTTCTGGATTTACACGCTAT......GTATATTTACAAGT 1621

     170 eLeuSerLys...MetGlyGlyTrpLeuLeuProGlyLysAspGlyAsnT 186
         :   ::::::   :::||||||||   :::|||   ::::::::::::
   1622 AACTGGCCGTGACATCGGTGGTTGGACAGTACCT...CAAGAATCAGCTG 1668

     186 hrValLysLeuAspValThrSerLeuLys...AsnAspLeu......... 198
           :::|||::::::|||:::      |||   :::||||||
   1669 GAACAAAAATTTCAGTTTCTCAAGCAAAAGCTGGAGACTTATTATTCTGG 1718

     199 AsnSerLeuValAsnLysTyrAsn.........GlnIleAsnSerAsnTh 212
         :::|||      :::::::|||:::          ::::::::::::
   1719 GGTTCTGCTGGCGGAACTTATCACGTAGCTATTTCATTAGGTGGCGGACA 1768

     212 rValLeuProAlaGlnSerGlySerGlyValLysValAlaThrGluA 229
         :::   ::::::|||:::|||   :::||||||||||::::::
   1769 ATATATCCATGCACCACAACCAGGTGAAAATGTTAAAGTTGGTTCAGTT. 1817

     229 laGluAlaArgGlnTrpLeuSerGluLeuAsnLeuProAsnSerCysLeu 245
               ||||||
   1818 ...........CAATGG.............................. 1823

     246 LysSerTyrGlySerGlyTyrValValThrVal 256
           |||   ::::::::::::|||::::::
   1824 ......TATACACCTGATTTTGCAGTAAGTATG 1850

seq_name: gb_in1:DMPUMI

seq_documentation_block:
LOCUS       DMPUMI       6120 bp    mRNA            INV       09-JUN-1992
DEFINITION  D.melanogaster mRNA for pumilio protein.
```

```
          |||:::|||||||:::                 :::::::::::::::::
    4610 CCGCTATATCCAGCAAAT..............AATACTGTTAGTCAGGA 4644

     229 aGluAlaArgGlnTrpLeuSerGluLeuAsnLeuProAsnSerCysLeuL 246
         ::::|||   :::|||||||:::|||||||:::   :::   :::   :::
    4645 ACAAGCAAATAAATGGCTTACAGAATTAGGTGGAACAATCGGCAAGGTAT 4694

     246 ysSerTyrGlySerGlyTyrValValThrValAspLeuThrProLeuGln 262
              ::::::|||||||||||||:::::::::::::||||||::::::
    4695 CTCAAAAAAACGGGGGATATGTTGTCAGTATAAACATGACCCCAATAGAC 4744

     263 LysMetVal 265
         :::|||:::
    4745 AATATGTTA 4753

seq_name: gb_ba1:ROPEPC

seq_documentation_block:
LOCUS       ROPEPC       2907 bp    DNA             BCT       02-JUN-1998
DEFINITION  R.obamensis PEPC gene.
ACCESSION   X99379
NID         g1435177
VERSION     X99379.1  GI:1435177
KEYWORDS    PEPC gene; phosphoenolpyruvate carboxylase.
SOURCE      Rhodothermus obamensis.
  ORGANISM  Rhodothermus obamensis
            Bacteria; Cytophagales; Rhodothermus.
REFERENCE   1  (bases 1 to 2907)
  AUTHORS   Takai,K., Sako,Y. and Uchida,A.
  TITLE     ppc, the gene for phosphoenolpyruvate carboxylase from an extremely
            thermophilic bacterium, Rhodothermus obamensis: cloning, sequencing
            and overexpression in Escherichia coli
  JOURNAL   Microbiology 144 (Pt 5), 1423-1434 (1998)
  MEDLINE   98274750
REFERENCE   2  (bases 1 to 2907)
  AUTHORS   Takai,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUL-1996) K. Takai, Laboratory of Marine
            Microbiology, Department of Applied Bioscience, Graduate School of
            Agriculture, Kyoto University, Kyoto 606-01, Japan
FEATURES             Location/Qualifiers
     source          1. .2907
                     /organism="Rhodothermus obamensis"
                     /strain="OKD7"
                     /db_xref="taxon:49320"
     gene            13. .2823
                     /gene="PEPC"
     CDS             13. .2823
                     /gene="PEPC"
                     /EC_number="4.1.1.31"
                     /codon_start=1
                     /transl_table=11
                     /product="phosphoenolpyruvate carboxylase"
                     /protein_id="CAA67760.1"
                     /db_xref="PID:e254301"
                     /db_xref="PID:g1435178"
                     /db_xref="GI:1435178"
                     /db_xref="SWISS-PROT:Q59757"
                     /translation="MLPPLQIEIEGTGISRPLSEHVNLLGGLLGQVIQEMAGPEMLEL
                     VETLRRLCKQAAQENRPEFREQAYTRIHSATYDELLWLLRAYTAFFHLVNQAEQQEII
                     RINRERAQQSTPERPRPESIDEAILALKQQGRTLDDVLTLLERLDIQPTVTAHPTEAR
                     RRSILYKQQHIAQMLSQQRRCQLTPEEQETLLLDLHNQITLLLGTAEVREERPTVRDE
                     VEQGLYFIQSTIWEAVPRIYEDVRRALRRYYGADVDFRPFLRYRSWIGSDRDGNPYVT
                     PEITRWTALTQRRLALQRYMEELRQLRRRLSLSDRYVAPPEELRRSLARDAREVSLPP
                     HVLRQFRHESFRLKISYIMGRLHGLLQALDDPTQPAPDYDADAFVEDLRLLQRCLEAC
                     GLERIARHDQLTRLLVLAQTFGFHLVTLDVRQHSSVHEAAVAELLRLAGVENDYRALP
                     ESRRQELLAEELSNPRPLLPPGARVSEATRQVLETFAVIRELVQLDPRLVGSYIVSMT
                     HTVSDLLEPMLLAKEVGLWHYERDPRTGKPGHVRCPIDFVPLFETIEDLEAAASRMEA
                     ILSHPVYRMQVAARGGFQEIMLGYSDSTKDGGYWMANWALHRAQEQLAEVCLRHGVDF
                     RLFHGRGGTVGRGGGRANQAILAMPPVVVHNGRIRFTEQGEVISFRYALPEIAHRHLEQ
                     IVNAMLRVVGLPAASGTDGTDPATRNRLMDELAARSMRAYRRLIDAPDFWSWYTRITP
                     IDQISRLPIASRPVSRSSAREVDFESLRAIPWVFAWTQVRYLIPGWFGIGQALDELLQ
                     TSPEHLETLRTWYRSWPFFRTVLQNAQREMVRARLEIAAYYDRLLGDGPTAFHQMIEE
                     DYHRARTAILRITDQESLLDHDPIIRKSVQLRNPYTDVLNLVQLELMRRIRSGAEADR
                     EPLRRALFLSINGIAAAMQSTG"
BASE COUNT      503 a    999 c    920 g    485 t
ORIGIN


alignment_scores:
            Quality:  104.00            Length:    315
              Ratio:  0.693               Gaps:     15
 Percent Similarity:  47.619   Percent Identity: 18.095

alignment_block:
 US-09-068-804-7 x ROPEPC   ..

 Align seg 1/1  to: ROPEPC  from: 1  to: 2907

     15 IleValAlaGluArgProGlnThrProSerAla...........SerGl 27
        |||      ||||||:::|||   :::::::::            :::||
    322 ATCAACCGGGAACGGGCACAGCAGAGTACGCCCGAACGTCCGCGTCCGGA 371

     27 uHisValGluThrAlaValValProSerThrThrGluHisArgGly...T 43
        |   ::::::   |||:::::::::   :::   :::   |||
    372 GTCGATCGACGAGGCCATCCTGGCCCTCAAGCAGCAGGGCCGGACGCTGG 421

     43 hrAspIleIleSerLeuSerGlnAlaAlaThrLysIleHisGlnAlaGln 59
          |||:::::::::|||
    422 ACGACGTGCTGACGCTG..........................CTG 441

     60 GlnThrLeuGlnSerThrProProIleSerGluGluAsnAsnAspGluAr 76
        :::   |||:::      |||:::::::::::::::   ::::::::::||
    442 GAGCGGCTGGACATCCAGCCCACGGTGACCGCCCATCCCACCGAGGCGCG 491

     76 g.........ThrLeuAlaArgGlnGlnLeuThrSerSerLeuAsnAlaL 90
        |    :::|||   :::||||||   ::::::   :::   :::
    492 CCGTCGCAGCATCCTCTACAAGCAGCAGCACATTGCCCAGATGCTTTCCC 541

     90 euAlaLysSerGlyValSerLeuSerAlaGluGlnAsnGluAsnLeuArg 106
           :::            |||:::::::|||:::::::|||:::|||
    542 AGCAACGCCGC...TGTCAGCTCACGCCCGAGGAGCAGGAGACCCTGCTG 588

    107 SerAlaPheSerAlaProThrSerAlaLeuPheSerAlaSerProMetAl 123
          ::::::   ::::::::::::   |||:::::::::::::   :::
    589 CTGGACCTGCACAATCAGATCACGTTGCTGCTGGGCACGGCCGAAGTGCG 638

    123 aGlnProArgThrThrIleSerAsp..................... 131
         :::   |||:::|||:::   |||
    639 CGAGGAGCGCCCCACCGTTCGCGACGAAGTCGAACAGGGCCTGTACTTTA 688

    132 .....AlaGluIleTrpAspMetValSerGlnAsnIleSerAlaIleGly 146
           :::   ||||||:::   |||:::::
    689 TCCAGAGTACGATCTGGGAGGCGGTGCCCCGC.................. 720

    147 AspSerTyrLeuGlyValTyrGluAsnValValAlaValTyrThrAspPh 163
                      :::||||||:::|||      :::         ::
    721 ..............ATCTATGAAGACGTGCGCCGGGCGCTGCGCCGCTA 755

    163 eTyrGlnAla.........PheSerAspIleLeuSerLysMetGlyGlyT 177
        :|||   |||         |||      :::|||   :::      :::|
    756 CTACGGCGCCGATGTGGACTTCCGGCCGTTTTTG...CGCTACCGCTCCT 802

    177 rpLeuLeuProGlyLysAspGlyAsnThr..................... 186
        ||:::   :::::::::|||||||||:::
    803 GGATCGGGAGCGACCGCGACGGCAACCCGTACGTCACGCCGGAAATCACG 852

    187 .....................ValLysLeuAspValThrSerLeuLysAs 196
                                :::|||::::::         :::::
    853 CGCTGGACGGCCCTGACGCAACGACGTTTGGCGCTCCAGCGCTACATGGA 902

    196 nAspLeuAsn.................SerLeuValAsnLysTyrAsnG 207
        :::::|||                   ||||||   :::::::|||
    903 AGAGCTGCGTCAGCTCCGGCGTCGGTTGAGCCTGTCGGATCGTTAC.... 948

    207 lnIleAsnSerAsnThrValLeuPheProAlaGlnSerGlySerGlyVal 223
```

```
                NQLFCWMSQERTTYVSSMINRSIDEMAIHNGVVLTSDNKKNIFAAIEKKFPDIKLDEK
                SAQTSISHTALNEIASSGLRAKILKRYSSDMDLFNTQMKDLTNLVSSSVYDKIFNEST
                KVLQIEISAEVLKAVYRQSNTN"
-10_signal      556. .561
                /gene="ippI"
gene            556. .1073
                /gene="ippI"
RBS             595. .600
                /gene="ippI"
CDS             606. .1073
                /gene="ippI"
                /codon_start=1
                /transl_table=11
                /product="antigen"
                /protein_id="AAA98423.1"
                /db_xref="PID:g155329"
                /db_xref="GI:155329"
                /translation="MSLNITENESISTAVIDAINSGATLKDINAIPDDMMDDIYSYAY
                DFYNKGRIEEAEVFFRFLCIYDFYNVDYIMGLAAIYQIKEQFQQAADLYAVAFALGKN
                DYTPVFHTGQCQLRLKAPLKAKECFELVIQHSNDEKLKIKAQSYLDAIQDIKE"
gene            1065. .2821
                /gene="ipaB"
RBS             1065. .1070
                /gene="ipaB"
CDS             1079. .2821
                /gene="ipaB"
                /codon_start=1
                /transl_table=11
                /product="antigen"
                /protein_id="AAA98424.1"
                /db_xref="PID:g155330"
                /db_xref="GI:155330"
                /translation="MHNVSTTTTGFPLAKILTSTELGDNTIQAANDAANKLFSLTIAD
                LTANQNINTTNAHSTSNILIPELKAPKSLNASSQLTLLIGNLIQILGEKSLTALTNKI
                TAWKSQQQARQQKNLEFSDKINTLLSETEGLTRDYEKQINKLKNADSKIKDLENKINQ
                IQTRLSELDPESPEKKKLSREEIQLTIKKDAAVKDRTLIEQKTLSIHSKLTDKSMQLE
                KEIDSFSAFSNTASAEQLSTQQKSLTGLASVTQLMATFIQLVGKNNEESLKNDLALFQ
                SLQESRKTEMERKSDEYAAEVRKAEELNRVMGCVGKILGALLTIVSVVAAAFSGGASL
                ALAAVGLALMVTDAIVQAATGNSFMEQALNPIMKAVIEPLIKLLSDAFTKMLEGLGVD
                SKKAKMIGSILGAIAGALVLVAAVVLVATVGKQAAAKLAENIGKIIGKTLTDLIPKFL
                KNFSSQLDDLITNAVARLNKFLGAAGDEVISKQIISTHLNQAVLLGESVNSATQAGGS
                VASAVFQNSASTNLADLTLSKYQVEQLSKYISEAIEKFGQLQEVIADLLASMSNSQAN
                RTDVAKAILQQTTA"
gene            2723. .3932
                /gene="ipaC"
-35_signal      2723. .2728
                /gene="ipaC"
-10_signal      2745. .2750
                /gene="ipaC"
RBS             2829. .2834
                /gene="ipaC"
CDS             2841. .3932
                /gene="ipaC"
                /codon_start=1
                /transl_table=11
                /product="antigen"
                /protein_id="AAA98425.1"
                /db_xref="PID:g155331"
                /db_xref="GI:155331"
                /translation="MEIQNTKPTQTLYTDISTKQTQSSSETQKSQNYQQIAAHIPLNV
                GKNPVLTTTLNDDQLLKLSEQVQHDSEIIARLTDKKMKDLSEMSHTLTPENTLDISSL
                SSNAVSLIISVAVLLSALRTAETKLGSQLSLIAFDATKSAAENIVRQGLAALSSSITG
                AVTQVGITGIGAKKTHSGISDQKGALRKNLATAQSLEKELAGSKLGLNKQIDTNITSP
                QTNSSTKFLGKNKLAPDNISLSTEHKTSLSSPDISLQDKIDTQRRTYELNTLSAQQKQ
                NIGRATMETSAVAGNISTSGGRYASALEEEEQLISQASSKQAEEASQVSKEASQATNQ
                LIQKLLNIIDSINQSKNSAASQIAGNIRA"
RBS             3971. .3977
                /gene="ipaD"
gene            3971. .4761
                /gene="ipaD"
CDS             3983. .>4761
                /gene="ipaD"
                /codon_start=1
                /transl_table=11
                /product="antigen"
                /protein_id="AAA98426.1"
                /db_xref="PID:g552045"
                /db_xref="GI:552045"
                /translation="MNITTLTNSISTSSFSPNNTNGSSTETVNSDIKTTTSSHPVSSL
                TMLNDTLHNIRTTNQALKKELSQKTLTKTSLEEIALHSSQISMDVNKSAQLLDILSRN
                EYPINKDARELLHSAPKEAELDGDQMISHRELWAKIANSINDINEQYLKVYEHAVSSY
                TQMYQDFSAVLSSLAGWISPGGNDGNSVKLQVNSLKKALEELKEKYKDKPLYPANNTV
                SQEQANKWLTELGGTIGKVSQKNGGYVVSINMTPIDNMLKS"
BASE COUNT     1769 a    952 c    805 g   1235 t
ORIGIN


alignment_scores:
           Quality:  302.50              Length:    253
             Ratio:  1.690                 Gaps:      7
 Percent Similarity:  70.751   Percent Identity: 29.644

alignment_block:
 US-09-068-804-7 x W10PPIBCA   ..

 Align seg 1/1  to: W10PPIBCA  from: 1  to: 4761

      24 SerAlaSerGluHisValGluThrAlaValValProSerThrThrGluHi 40
         |||::::::|||   |||:::::::::::::   ::::::|||:::   ||
    4049 TCATCAACCGAAACAGTTAATTCTGATATAAAAACAACGACCAGTTCTCA 4098

      40 sArgGlyThrAspIleIleSerLeuSerGlnAlaAlaThrLysIleHisG 57
         |      ::::::::::::   |||:::::::::      ::::|||:::
    4099 TCCTGTAAGTTCCCTTACTATGCTCAACGACACCCTTCATAATATCAGAA 4148

      57 lnAlaGlnGlnThrLeuGln.........SerThrProProIleSerGlu 70
           ::::::|||:::|||:::          |||      :::::::::
    4149 CAACAAATCAGGCATTAAAGAAAGAGCTTTCACAAAAAACGTTGACTAAA 4198

      71 GluAsnAsnAspGluArgThrLeuAlaArgGlnGlnLeuThrSerSer.. 86
            :::   :::|||   :::|||         |||::::::   :::
    4199 ACATCGCTAGAAGAAATAGCATTACATTCATCTCAGATTAGCATGGATGT 4248

      87 ...................LeuAsnAlaLeuAlaLysSerGlyValSerL 97
                             |||:::   |||::::::::      ::::
    4249 AAATAAATCCGCTCAACTATTGGATATTCTTTCCAGGAACGAATATCCAA 4298

      97 euSerAlaGluGlnAsnGluAsnLeuArgSerAlaPheSerAlaProThr 113
         :::::   :::      |||   |||:::         |||||||||:::
    4299 TTAATAAAGACGCAAGAGAATTATTACAT.........TCAGCCCCGAAA 4339

     114 SerAlaLeuPheSerAlaSerProMetAlaGlnProArgThrThrIleSe 130
            |||   ::::::::::::::::|||                |||||
    4340 GAAGCCGAGCTTGATGGAGATCAAATG..................ATATC 4371

     130 rAspAlaGluIleTrpAspMetValSerGlnAsnIleSerAlaIleGlyA 147
         |:::   |||:::|||:::::::::::::::::::|||:::::|||::::
    4372 TCATAGAGAACTGTGGGCTAAAATTGCAAACTCCATCAATGATATTAATG 4421

     147 spSerTyrLeuGlyValTyrGluAsnValValAlaValTyrThrAspPhe 163
         ::   ||||||   ||||||||:::::::|||:::   ||||||:::
    4422 AACAGTATCTGAAAGTATATGAACATGCCGTTAGTTCATATACTCAAATG 4471

     164 TyrGlnAlaPheSerAspIleLeuSerLysMetGlyGlyTrpLeuLeuPr 180
         ||||||:::|||||:::::::||||||   ::::::||||||:::   ||
    4472 TATCAAGATTTTAGCGCTGTTCTTTCCAGTCTTGCCGGCTGGATCTCTCC 4521

     180 o...GlyLysAspGlyAsnThrValLysLeuAspValThrSerLeuLysA 196
         |   |||:::|||||||||:::||||||||||:::|||:::||||||||:
    4522 CGGAGGTAACGACGGAAACTCCGTGAAATTACAAGTCAACTCGCTTAAAA 4571

     196 snAspLeuAsnSerLeuValAsnLysTyrAsnGlnIleAsnSerAsnThr 212
         :::::|||:::   |||   :::||||||:::            ::::::
    4572 AGGCATTGGAAGAACTCAAGGAAAAATATAAA............GATAAA 4609

     213 ValLeuPheProAlaGlnSerGlySerGlyValLysValAlaThrGluAl 229
```

```
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA33383.1"
                     /db_xref="PID:g47057"
                     /db_xref="GI:47057"
                     /db_xref="SWISS-PROT:P18013"
                     /translation="MNITTLTNSISTSSFSPNNTNGSSTETVNSDIKTTTSSHPVSSL
                     TMLNDTLHNIRTTNQALKKELSQKTLTKTSLEEIALHSSQISMDVNKSAQLLDILSRH
                     EYPINKDARELLHSAPKEAELDGDQMISHRELWAKIANSINDINEQYLKVYEHAVSSY
                     TQMYQDFSAVLSSLAGWISPGGNDGNSVKLQVNSLKKALEELKEKYKDKPLYPANNTV
                     SQEQANKWLTELGGTIGKVSQKNGGYVVSINMTPIDNMLKSLDNLGGNGEVVLDNAKY
                     QAWNAGFSAEDETMKNNLQTLVQKYSNANSIFDNLVKVLSSTISSCTDTDKLFLHF"
BASE COUNT      2845 a    1501 c    1246 g    2071 t
ORIGIN


alignment_scores:
            Quality:  506.00              Length:     330
              Ratio:  2.049                 Gaps:       8
 Percent Similarity:  74.848    Percent Identity: 36.667

alignment_block:
 US-09-068-804-7 x SFPVR2   ..

 Align seg 1/1  to: SFPVR2  from: 1  to: 7663

      24 SerAlaSerGluHisValGluThrAlaValValProSerThrThrGluHi 40
         |||:::::::|||   |||:::::::::::   :::::::|||:::   ||
    5496 TCATCAACCGAAACAGTTAATTCTGATATAAAAACAACGACCAGTTCTCA 5545

      40 sArgGlyThrAspIleIleSerLeuSerGlnAlaAlaThrLysIleHisG 57
         |     ::::::::::::   |||:::::::::      :::|||:::
    5546 TCCTGTAAGTTCCCTTACTATGCTCAACGACACCCTTCATAATATCAGAA 5595

      57 lnAlaGlnGlnThrLeuGln.........SerThrProProIleSerGlu 70
           ::::::|||:::|||:::         |||      :::::::::
    5596 CAACAAATCAGGCATTAAAGAAAGAGCTTTCACAAAAAACGTTGACTAAA 5645

      71 GluAsnAsnAspGluArgThrLeuAlaArgGlnGlnLeuThrSerSer.. 86
         :::   :::|||   :::|||         |||::::::   :::
    5646 ACATCGCTAGAAGAAATAGCATTACATTCATCTCAGATTAGCATGGATGT 5695

      87 ...................LeuAsnAlaLeuAlaLysSerGlyValSerL 97
                            |||:::   |||::::::         ::::
    5696 AAATAAATCCGCTCAACTATTGGATATTCTTTCCAGGCACGAATATCCAA 5745

      97 euSerAlaGluGlnAsnGluAsnLeuArgSerAlaPheSerAlaProThr 113
         :::::   :::      |||   |||:::         |||||||||:::
    5746 TTAATAAAGACGCAAGAGAATTATTACAT.........TCAGCTCCGAAA 5786

     114 SerAlaLeuPheSerAlaSerProMetAlaGlnProArgThrThrIleSe 130
         |||   ::::::::::::::|||                  |||||
    5787 GAAGCCGAGCTTGATGGAGATCAAATG..................ATATC 5818

     130 rAspAlaGluIleTrpAspMetValSerGlnAsnIleSerAlaIleGlyA 147
         |:::   |||:::|||:::::::::::::::::::::|||:::::::|||:::
    5819 TCATAGAGAACTGTGGGCTAAAATTGCAAACTCCATCAATGATATTAATG 5868

     147 spSerTyrLeuGlyValTyrGluAsnValValAlaValTyrThrAspPhe 163
         ::   ||||||   |||||||||:::::|||:::   ||||||:::
    5869 AACAGTATCTGAAAGTATATGAACATGCCGTTAGTTCATATACTCAAATG 5918

     164 TyrGlnAlaPheSerAspIleLeuSerLysMetGlyGlyTrpLeuLeuPr 180
         ||||||:::||||||:::::||||||   ::::::||||||:::   ||
    5919 TATCAAGATTTTAGCGCTGTTCTTTCCAGTCTTGCCGGCTGGATCTCTCC 5968

     180 o...GlyLysAspGlyAsnThrValLysLeuAspValThrSerLeuLysA 196
         |   |||:::|||||||||:::|||||||||:::|||:::|||||||||:
    5969 CGGAGGTAACGACGGAAACTCCGTGAAATTACAAGTCAACTCGCTTAAAA 6018

     196 snAspLeuAsnSerLeuValAsnLysTyrAsnGlnIleAsnSerAsnThr 212
         :::::|||:::   |||   :::|||||||:::            ::::::
    6019 AGGCATTGGAAGAACTCAAGGAAAAATATAAA.............GATAAA 6056

     213 ValLeuPheProAlaGlnSerGlySerGlyValLysValAlaThrGluAl 229
            |||:::||||||:::                 ::::::::::::::::
    6057 CCGCTATATCCAGCAAAT..............AATACTGTTAGTCAGGA 6091

     229 aGluAlaArgGlnTrpLeuSerGluLeuAsnLeuProAsnSerCysLeuL 246
         ::::|||   :::||||||:::||||||:::   :::   :::   :::
    6092 ACAAGCAAATAAATGGCTTACAGAATTAGGTGGAACAATCGGCAAGGTAT 6141

     246 ysSerTyrGlySerGlyTyrValValThrValAspLeuThrProLeuGln 262
                :::::|||||||||||||:::::::::::|||||||:::::
    6142 CTCAAAAAAACGGGGGATATGTTGTCAGTATAAACATGACCCCAATAGAC 6191

     263 LysMetValGlnAspIleAspGlyLeuGlyAlaProGlyLysAspSerLy 279
         :::|||:::::::::::::|||:::||||||:::         ::::::
    6192 AATATGTTAAAAAGCTTAGATAATCTAGGTGGA.........AATGGCGA 6232

     279 sLeuGluMetAspAsnAlaLysTyrGlnAlaTrpGlnSerGlyPheLysA 296
          :::   :::|||||||||||||||||||||||||:::::|||||   |
    6233 GGTTGTGCTAGATAATGCAAAATATCAGGCATGGAATGCCGGATTCTCTG 6282

     296 laGlnGluGluAsnMetLysThrThrLeuGlnThrLeuThrGlnLysTyr 312
         ||:::::|||:::||||||:::::||||||||||||:::||||||||||
    6283 CCGAAGATGAAACAATGAAAAATAATCTTCAAACTTTAGTTCAAAAATAC 6332

     313 SerAsnAlaAsnSerLeuTyrAspAsnLeuValLysValLeuSerSerTh 329
         ||||||||||||||:::::|||||||||||||||||||||||||||||||
    6333 AGTAATGCCAATAGTATTTTTGATAATTTAGTAAAGGTTTTGAGTAGTAC 6382

     329 rIleSerSerSerLeuGluThrAlaLysSerPheLeuGln 342
         ||||||||||:::   :::|||:::|||   ||||||:::
    6383 AATAAGCTCATGTACAGATACAGATAAACTTTTTCTCCAT 6422

seq_name: gb_ba1:W10PPIBCA

seq_documentation_block:
LOCUS       W10PPIBCA    4761 bp    DNA             BCT       07-MAY-1996
DEFINITION  Plasmid pWR100 (from S.flexneri) antigen ippI, ipaB, ipaC genes,
            complete cds, and ippD gene, 3' end.
ACCESSION   M34849
NID         g155327
VERSION     M34849.1  GI:155327
KEYWORDS    antigen.
SOURCE      Plasmid pWR100 (clone: HB101/pHS[4108,5357].).
  ORGANISM  Plasmid pWR100
            plasmids.
REFERENCE   1  (bases 1 to 4761; 1 to 538; 3987 to 4761)
  AUTHORS   Baudry,B., Kaczorek,M. and Sansonetti,P.J.
  TITLE     Nucleotide sequence of the invasion plasmid antigen B and C genes
            (ipaB and ipaC) of Shigella flexneri
  JOURNAL   Microb. Pathog. 4 (5), 345-357 (1988)
  MEDLINE   89200844
COMMENT     Draft entry and computer-readable sequence for [1] kindly submitted
            by B.Baudry, 05-JUN-1990.
               Center for Vaccine Development
               University of Maryland at Baltimore
               10 South Pine Street
               Baltimore, MD 21201.
FEATURES             Location/Qualifiers
     source          1. .4761
                     /organism="Plasmid pWR100"
                     /plasmid="Plasmid pWR100"
                     /specific_host="M90T"
                     /db_xref="taxon:2661"
                     /clone="HB101/pHS[4108,5357]."
     CDS             <1. .549
                     /note="ORF"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA98422.1"
                     /db_xref="PID:g155328"
                     /db_xref="GI:155328"
                     /translation="KLVKKILSVFNLKQRFPQKNFGCPVNINKIRDSVIDKIKDSNSG
```

```
 4966 ACGCATTGACAACACTCAAGAAAAACTATGAA...........GATAAA 5003

  213 ValLeuPheProAlaGlnSerGlySerGlyValLysValAlaThrGluAl 229
      |||:::||||||               ::::::::::::::|||
 5004 CCGCTATATCCAGCA.............ACTAATACTGTTAGTGAGCA 5038

  229 aGluAlaArgGlnTrpLeuSerGluLeuAsnLeuProAsnSerCysLeuL 246
      ||||||   :::|||||||:::||||||:::   :::   :::   :::
 5039 AGAAGCAAATAAATGGCTTACAGAATTAGGTGGAACAATCGGCACGGTAT 5088

  246 ysSerTyrGlySerGlyTyrValValThrValAspLeuThrProLeuGln 262
       :::   ::::::||||||||||||||||||||||||||||||||:::
 5089 CTGCAAAAAACGGGGGATATGTTGTCAGTATAAACATGACCCCAATATAT 5138

  263 LysMetValGlnAspIleAspGlyLeuGlyAlaProGlyLysAspSerLy 279
      :::|||::::::   :::|||:::||||||:::         ::::::
 5139 AATATGTTAAACCGCTTAGATAATCTAGGTGGA.........AATGGCGA 5179

  279 sLeuGluMetAspAsnAlaLysTyrGlnAlaTrpGlnSerGlyPheLysA 296
       :::   :::||||||||||||||||||||||||||:::::||||||   |
 5180 GGTTGTGCTAGATAATGCAAAATATCAGGCATGGAATGCCGGATTCTCTG 5229

  296 laGlnGluGluAsnMetLysThrThrLeuGlnThrLeuThrGlnLysTyr 312
      ||:::::::|||:::||||||:::::|||||||||||||:::||||||||
 5230 CCGAAGATGAAACAATGAAAAATAATCTTCAAACTTTAGTTCAAAAATAC 5279

  313 SerAsnAlaAsnSerLeuTyrAspAsnLeuValLysValLeuSerSerTh 329
      ||||||||||||||||:::::|||||||||||||||||||||||||||||
 5280 AGTAATGCCAATAGTATTTTTGATAATTTAGTAAAGGTTTTGAGTAGTAC 5329

  329 rIleSerSerSerLeuGluThrAlaLysSerPheLeuGln 342
      ||||||||||:::   :::|||:::|||   ||||||:::
 5330 AATAAGCTCATGTACAGATACAGATAAACTTTTTCTCCAT 5369
```

seq_name: gb_ba1:SFPVR2

seq_documentation_block:

```
LOCUS       SFPVR2       7663 bp    DNA             BCT       12-SEP-1993
DEFINITION  Shigella flexineri plasmid YSH6000 virulence region-2 DNA
            containing 24k gene, 18k gene, ipaB gene, ipaC gene and ipaD gene.
ACCESSION   X15319
NID         g47052
VERSION     X15319.1  GI:47052
KEYWORDS    ipaB gene; ipaC gene; ipaD gene; virulence proteins.
SOURCE      Enterobacteriaceae.
  ORGANISM  Enterobacteriaceae
            Bacteria; Proteobacteria; gamma subdivision.
REFERENCE   1  (bases 1 to 7663)
  AUTHORS   Sasakawa,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAY-1989) Sasakawa C., Institute of Medical Science,
            University of Tokyo, Shirokanedai 4-6-1, Minato-ku Tokyo 108, Japan
REFERENCE   2  (bases 1010 to 6538)
  AUTHORS   Sasakawa,C., Adler,B., Tobe,T., Okada,N., Nagai,S., Komatsu,K. and
            Yoshikawa,M.
  TITLE     Functional organization and nucleotide sequence of virulence
            Region-2 on the large virulence plasmid in Shigella flexneri 2a
  JOURNAL   Mol. Microbiol. 3 (9), 1191-1201 (1989)
  MEDLINE   90014179
COMMENT     Region 2 has three promoters that transcribe RNAs of 4.0, 4.5 and
            7.5kb. The 4.0 kb transcript produces the 24,18kD, IpaB and IpaC
            proteins. The 4.5 kb transcript produces the ipaD protein. The
            7.5kb transcript covers the entire region 2.
            Data kindly reviewed (20-NOV-1989) by Sasakawa C.
FEATURES             Location/Qualifiers
     source          1. .7663
                     /organism="Enterobacteriaceae"
                     /plasmid="YSH6000"
                     /db_xref="taxon:543"
     promoter        1291. .1296
                     /note="-35 region for 4kb transcript"
     promoter        1312. .1318
                     /note="-10 region for 4kb transcript"
     RBS             1360. .1362
                     /note="put. ribosome binding site"
     CDS             1370. .1996
                     /note="24k gene (AA 1-208)"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA33379.1"
                     /db_xref="PID:g47053"
                     /db_xref="GI:47053"
                     /db_xref="SWISS-PROT:P33548"
                     /translation="MQILNKILPQVEFAIPRPSFDSLSRNKLVKKILSVFNLKQRFPQ
                     KNFGCPVNINKIRDSVIDKIKDSNSGNQLFCWMSQERTTYVSSMINRSIDEMAIHNGV
                     VLTSDNKKNIFAAIEKKFPDIKLDEKSAQTSISHTALNEIASSGLRAKILKRYSSDMD
                     LFNTQMKDLTNLVSSSVYDKIFNESTKVLQIEISAEVLKAVYRQSNTN"
     RBS             2043. .2045
                     /note="put. ribosome binding site"
     CDS             2053. .2520
                     /note="18k gene (AA 1-155)"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA33380.1"
                     /db_xref="PID:g47054"
                     /db_xref="GI:47054"
                     /db_xref="SWISS-PROT:P18008"
                     /translation="MSLNITENESISTAVIDAINSGATLKDINAIPDDMMDDIYSYAY
                     DFYNKGRIEEAEVFFRFLCIYDFYNVDYIMGLAAIYQIKEQFQQAADLYAVAFALGKN
                     DYTPVFHTGQCQLRLKAPLKAKECFELVIQHSNDEKLKIKAQSYLDAIQDIKE"
     RBS             2513. .2515
                     /note="put. ribosome binding site"
     CDS             2526. .4268
                     /note="ipaB gene (AA 1-580)"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA33381.1"
                     /db_xref="PID:g47055"
                     /db_xref="GI:47055"
                     /db_xref="SWISS-PROT:P18011"
                     /translation="MHNVSTTTTGFPLAKILASTELGDNTIQAANDAANKLFSLTIAD
                     LTANQNINTTNAHSTSNILIPELKAPKSLNASSQLTLLIGNLIQILGEKSLTALTNKI
                     TAWKSQQQARQQKNLEFSDKINTLLSETEGLTRDYEKQINKLKNADSKIKDLENKINQ
                     IQTRLSELDPESPEKKKLSREEIQLTIKKDAAVKDRTLIEQKTLSIHSKLTDKSMQLE
                     KEIDSFSAFSNTASAEQLSTQQKSLTGLASVTQLMATFIQLVGKNNEESLKNDLALFQ
                     SLQESRKTEMERKSDEYAAEVRKAEELNRVMGCVGKILGALLTIVSVVAAAFSGGASL
                     ALAAVGLALMVTDAIVQAATGNSFMEQALNPIMKAVIEPLIKLLSDAFTKMLEGLGVD
                     SKKAKMIGSILGAIAGALVLVAAVVLVATVGKQAAAKLAENIGKIIGKTLTDLIPKFL
                     KNFSSQLDDLITNAVARLNKFLGAAGDEVISKQIISTHLNQAVLLGESVNSATQAGGS
                     VASAVFQNSASTNLADLTLSKYQVEQLSKYISEAIEKFGQLQEVIADLLASMSNSQAN
                     RTDVAKAILQQTTA"
     RBS             4224. .4226
                     /note="put. ribosome binding site"
     CDS             4231. .5379
                     /note="ipaC gene (AA 1-382)"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA33382.1"
                     /db_xref="PID:g47056"
                     /db_xref="GI:47056"
                     /db_xref="SWISS-PROT:P18012"
                     /translation="MLQKQFCNKLLLDTNKENVMEIQNTKPTQILYTDISTKQTQSSS
                     ETQKSQNYQQIAAHIPLNVGKNPVLTTTLNDDQLLKLSEQVQHDSEIIARLTDKKMKD
                     LSEMSHTLTPENTLDISSLSSNAVSLIISVAVLLSALRTAETKLGSQLSLIAFDATKS
                     AAENIVRQGLAALSSSITGAVTQVGITGIGAKKTHSGISDQKGALRKNLATAQSLEKE
                     LAGSKLGLNKQIDTNITSPQTNSSTKFLGKNKLAPDNISLSTEHKTSLSSPDISLQDK
                     IDTQRRTYELNTLSAQQKQNIGRATMETSAVAGNISTSGGRYASALEEEEQLISQASS
                     KQAEEASQVSKEASQATNQLIQKLLNIIDSINQSKNSTASQIAGNIRA"
     promoter        5376. .5381
                     /note="-35 region for 4.5kb transcript"
     promoter        5396. .5401
                     /note="-10 region for 4.5kb transcript"
     RBS             5421. .5423
                     /note="put. ribosome binding site"
     CDS             5430. .6428
                     /note="ipaD gene (AA 1-332)"
```

```
  AUTHORS   Palchaudhuri,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUN-1991) S. Palchaudhuri, Wayne State Univers.School
            of Medicine, Department of Immunology and, Microbiology, 540 E.
            Canfield, Detroit, MI 48201, USA
REFERENCE   2  (bases 1 to 6869)
  AUTHORS   Yao,R. and Palchaudhuri,S.
  TITLE     Nucleotide sequence of the ipaBCD structure genes of Shigella
            dysenteriae
  JOURNAL   Unpublished
FEATURES             Location/Qualifiers
     source          1. .6869
                     /organism="Shigella dysenteriae"
                     /strain="CG097"
                     /db_xref="taxon:622"
     CDS             317. .943
                     /note="24kD ORF"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA43188.1"
                     /db_xref="PID:g46933"
                     /db_xref="GI:46933"
                     /db_xref="SWISS-PROT:Q04314"
                     /translation="MQILNKILPQVEFAIPRPSFNSLSRNKLVKKILSVFNLKQRFPQ
                     KNFGCTVNINKIRDNVIDKIKDSNSGNQLFCWMSQERTTYVSSMINRSIDEMAIHNGV
                     VLTSDNKKNIFAAIEKKFPDIKLDEKSAQTSISHTALNEIASSGLRAKILKRYSSDMD
                     LFNTQMKDLTNLVSSSVYDKIFNESTKVLQIEISAEVLKAVYRQSNTN"
     CDS             1000. .1467
                     /note="IppI ORF"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA43189.1"
                     /db_xref="PID:g46934"
                     /db_xref="GI:46934"
                     /db_xref="SWISS-PROT:P18008"
                     /translation="MSLNITENESISTAVIDAINSGATLKDINAIPDDMMDDIYSYAY
                     DFYNKGRIEEAEVFFRFLCIYDFYNVDYIMGLAAIYQIKEQFQQAADLYAVAFALGKN
                     DYTPVFHTGQCQLRLKAPLKAKECFELVIQHSNDEKLKIKAQSYLDAIQDIKE"
     CDS             1473. .3215
                     /note="IpaB ORF"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA43190.1"
                     /db_xref="PID:g46935"
                     /db_xref="GI:46935"
                     /db_xref="SWISS-PROT:Q03945"
                     /translation="MHNVNTTTTGLSLAKILASTELGDNTIQAGNDAANKLFSLTIAD
                     LTANKNINTTNAHSTSNILIPELKAPKSLNASSQLTLLIGNLIQILGEKSLTALTNKI
                     TAWKSQQQARQQKNLEFSDKINTLLSETEGLTRDYEKQINKLKNADSKIKDLENKINQ
                     IQTRLSELDPDSPEKKKLSREEIQLTIKKDAAVKDRTLIEQKTLSIHSKLTDKSMQLE
                     KEIDSFSAFSNTASAEQLSTQQKSLTGLASVTQLMATFIQLVGKNNEESLKNDLALFQ
                     SLQESRKTEMERKSDEYAAEVRKAEELNRVMGCVGKILGALLTIVSVVAAAFSGGASL
                     ALADVGLALMVTDAIVQAATGNSFMEQALNPIMKAVIEPLIKLLSDAFTKMLEGLGVD
                     SKKAKMIGSILGAIAGALVLVAAVVLVATVGKQAAAKLAENIGKIIGKTLTDLIPKFL
                     KNFSSQLDDLITNAVARLNKFLGAAGDEVISKQIISTHLNQAVLLGESVNSATQAGGS
                     VASAVFQNSASTNLADLTLSKYQVEQLSKYISEAIEKFGQLQEVIADLLASMSNSQAN
                     RTDVAKAILQQTTA"
     CDS             3178. .4326
                     /note="IpaC ORF"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA43191.1"
                     /db_xref="PID:g46936"
                     /db_xref="GI:46936"
                     /db_xref="SWISS-PROT:Q03946"
                     /translation="MLQKQFCNKLLLDTNKENVMEIQNTKSTQILYTDISTKQTQSSS
                     ETQKSQNYQQIAAHIPLNVGKNPVLTTTLNDDQLLKLSEQVQHDSEIIARLTDKKMKD
                     LSEMSHTLTPENTLDISSLSSNAVSLIISVAVLLSALRTAETKLGSQLSLIAFDATKS
                     AAENIVRQGLAALSSSITGAVTQVGITGIGAKKTHSGISDQKGALRKNLATAQSLEKE
                     LAGSKLGLNKQIDTNITSPQTNSSTKFLGKNKLAPDNISLSTEHKTSLSSPDISLQDK
                     IDTQRRTYELNTLSAQQKQNIGRATMETSAVAGNISTSGGRYASALEEEEQLISQASS
                     KQAEEASQVSKEASQATNQLIQKLLNIIDNINQSRSSTASQIAGNIRA"
     CDS             4377. .5375
                     /note="IpaD ORF"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA43192.1"
                     /db_xref="PID:g46937"
                     /db_xref="GI:46937"
                     /db_xref="SWISS-PROT:Q03947"
                     /translation="MNITTLTNSISTSSFSPNNTNGSSTETVNSDIKTTTSSHPVSSL
                     TMLNDTLHNIRTTNQALKKELSQKTLTKTSLEEIALHSSQISMDVNKSAQLLNILSKT
                     EYPINKDARELLHSAPKEAELDGYEMISHRELWAKIANSINDINEQYLKVYEHAVSSY
                     TQMYQEFSAVLSSLAGWISPGGNDGNSVKLQVKSLKDALTTLKKNYEDKPLYPATNTV
                     SEQEANKWLTELGGTIGTVSAKNGGYVVSINMTPIYNMLNRLDNLGGNGEVVLDNAKY
                     QAWNAGFSAEDETMKNNLQTLVQKYSNANSIFDNLVKVLSSTISSCTDTDKLFLHF"
BASE COUNT     2518 a   1369 c   1135 g   1847 t
ORIGIN


alignment_scores:
            Quality:  506.00              Length:    330
              Ratio:  2.065                 Gaps:      8
 Percent Similarity:  74.242   Percent Identity: 37.576

alignment_block:
 US-09-068-804-7 x SDIPABCD   ..

 Align seg 1/1  to: SDIPABCD  from: 1  to: 6869

      24 SerAlaSerGluHisValGluThrAlaValValProSerThrThrGluHi 40
         |||::::::|||   |||::::::::::::   ::::::|||:::   ||
    4443 TCATCAACCGAAACAGTTAATTCTGATATAAAAACAACGACCAGTTCTCA 4492

      40 sArgGlyThrAspIleIleSerLeuSerGlnAlaAlaThrLysIleHisG 57
         |      ::::::::::::   |||:::::::::      :::|||:::
    4493 TCCTGTAAGTTCCCTTACTATGCTCAACGACACCCTTCATAATATCAGAA 4542

      57 lnAlaGlnGlnThrLeuGln.........SerThrProProIleSerGlu 70
           ::::::|||:::|||:::         |||      :::::::::
    4543 CAACAAATCAGGCATTAAAGAAAGAGCTTTCACAAAAAACGTTGACTAAA 4592

      71 GluAsnAsnAspGluArgThrLeuAlaArgGlnGlnLeuThrSerSer.. 86
         :::   :::|||   :::|||        |||::::::   :::
    4593 ACATCGCTAGAAGAAATAGCATTACATTCATCTCAGATTAGCATGGATGT 4642

      87 ...................LeuAsnAlaLeuAlaLysSerGlyValSerL 97
                             ||||||   |||:::|||:::      ::::
    4643 AAACAAATCCGCTCAACTATTGAATATTCTTTCCAAGACAGAATATCCAA 4692

      97 euSerAlaGluGlnAsnGluAsnLeuArgSerAlaPheSerAlaProThr 113
         :::::   :::      |||   |||:::         ||||||||:::
    4693 TTAATAAAGACGCAAGAGAATTATTACAT.........TCAGCCCCGAAA 4733

     114 SerAlaLeuPheSerAlaSerProMetAlaGlnProArgThrThrIleSe 130
         |||   :::::::::      |||                 |||||
    4734 GAAGCTGAGCTTGATGGATATGAAATG.................ATATC 4765

     130 rAspAlaGluIleTrpAspMetValSerGlnAsnIleSerAlaIleGlyA 147
         |:::   |||:::|||:::::::::::::::::::|||:::::|||::::
    4766 TCATAGAGAACTGTGGGCTAAAATTGCAAACTCCATCAATGATATTAATG 4815

     147 spSerTyrLeuGlyValTyrGluAsnValValAlaValTyrThrAspPhe 163
         ::   ||||||   |||||||||:::::|||:::   |||||||:::
    4816 AACAGTATCTGAAAGTATATGAACATGCCGTTAGTTCATATACTCAAATG 4865

     164 TyrGlnAlaPheSerAspIleLeuSerLysMetGlyGlyTrpLeuLeuPr 180
         ||||||:::||||||:::::||||||   ::::::||||||:::   ||
    4866 TATCAAGAATTTAGCGCTGTTCTTTCCAGTCTTGCCGGCTGGATCTCTCC 4915

     180 o...GlyLysAspGlyAsnThrValLysLeuAspValThrSerLeuLysA 196
         |   |||:::||||||||||:::||||||||||:::|||:::||||||||:
    4916 CGGAGGTAACGACGGAAACTCCGTGAAATTACAAGTCAAGTCGCTTAAAG 4965

     196 snAspLeuAsnSerLeuValAsnLysTyrAsnGlnIleAsnSerAsnThr 212
         :::::|||::::::|||   :::::::|||:::            ::::::
```

```
                VASAVFQNSASTNLADLTLSKYQVEQLSKYISEAIEKFGQLQEVIADLLASMSNSQAN
                RTDVAKAILQQTTA"
     CDS        2784. .3932
                /note="IpaC protein"
                /codon_start=1
                /transl_table=11
                /protein_id="AAA26523.1"
                /db_xref="PID:g152743"
                /db_xref="GI:152743"
                /translation="MLQKQFCNKLLLDTNKENVMEIQNTKPTQTLYTDISTKQTQSSS
                ETQKSQNYQQIAAHIPLNVGKNPVLTTTLNDDQLLKLSEQVQHDSEIIARLTDKKMKD
                LSEMSHTLTPENTLDISSLSSNAVSLIISVAVLLSALRTAETKLGSQLSLIAFDATKS
                AAENIVRQGLAALSSSITGAVTQVGITGIGAKKTHSGISDQKGALRKNLATAQSLEKE
                LAGSKLGLNKQIDTNITSPQTNSSTKFLGKNKLAPDNISLSTEHKTSLSSPDISLQDK
                IDTQRRTYELNTLSAQQKQNIGRATMETSAVAGNISTSGGRYASALEEEEQLISQASS
                KQAEEASQVSKEASQATNQLIQKLLNIIDSINQSKNSAASQIAGNIRA"
     CDS        3983. .4981
                /note="IpaD protein"
                /codon_start=1
                /transl_table=11
                /protein_id="AAA26524.1"
                /db_xref="PID:g152744"
                /db_xref="GI:152744"
                /translation="MNITTLTNSISTSSFSPNNTNGSSTETVNSDIKTTTSSHPVSSL
                TMLNDTLHNIRTTNQALKKELSQKTLTKTSLEEIALHSSQISMDVNKSAQLLDILSRN
                EYPINKDARELLHSAPKEAELDGDQMISHRELWAKIANSINDINEQYLKVYEHAVSSY
                TQMYQDFSAVLSSLAGWISPGGNDGNSVKLQVNSLKKALEELKEKYKDKPLYPANNTV
                SQEQANKWLTELGGTIGKVSQKNGGYVVSINMTPIDNMLKSLDNLGGNGEVVLDNAKY
                QAWNAGFSAEDETMKNNLQTLVQKYSNANSIFDNLVKVLSSTISSCTDTDKLFLHF"
     CDS        4990. .>5255
                /note="IpaA protein"
                /codon_start=1
                /transl_table=11
                /protein_id="AAA26525.1"
                /db_xref="PID:g551981"
                /db_xref="GI:551981"
                /translation="MHNVNNTQAPTFLYKATSPSSTEYSELKSKISDIHSSQTSLKTP
                ASVSEKENFATSFNQKCLDFLFSSSGKEDVLRSIYSNSMNAYAK"
BASE COUNT     1950 a    1037 c     889 g    1379 t
ORIGIN


alignment_scores:
           Quality:  508.00              Length:    330
             Ratio:  2.048                 Gaps:      8
 Percent Similarity:  75.152    Percent Identity: 36.667

alignment_block:
 US-09-068-804-7 x SHFIPABCD   ..

 Align seg 1/1  to: SHFIPABCD  from: 1  to: 5255

      24 SerAlaSerGluHisValGluThrAlaValValProSerThrThrGluHi 40
         |||::::::|||   |||:::::::::::::   ::::::|||:::   ||
    4049 TCATCAACCGAAACAGTTAATTCTGATATAAAAACAACGACCAGTTCTCA 4098

      40 sArgGlyThrAspIleIleSerLeuSerGlnAlaAlaThrLysIleHisG 57
         |      ::::::::::::   |||::::::::::      :::|||:::
    4099 TCCTGTAAGTTCCCTTACTATGCTCAACGACACCCTTCATAATATCAGAA 4148

      57 lnAlaGlnGlnThrLeuGln.........SerThrProProIleSerGlu 70
         ::::::|||:::|||:::          |||      :::::::::
    4149 CAACAAATCAGGCATTAAAGAAAGAGCTTTCACAAAAAACGTTGACTAAA 4198

      71 GluAsnAsnAspGluArgThrLeuAlaArgGlnGlnLeuThrSerSer.. 86
         :::   :::|||   :::|||        |||::::::   :::
    4199 ACATCGCTAGAAGAAATAGCATTACATTCATCTCAGATTAGCATGGATGT 4248

      87 ..................LeuAsnAlaLeuAlaLysSerGlyValSerL 97
                           |||:::   |||::::::::::      ::::
    4249 AAATAAATCCGCTCAACTATTGGATATTCTTTCCAGGAACGAATATCCAA 4298

      97 euSerAlaGluGlnAsnGluAsnLeuArgSerAlaPheSerAlaProThr 113
         :::::   :::      |||   |||:::         |||||||||:::
    4299 TTAATAAAGACGCAAGAGAATTATTACAT.........TCAGCCCCGAAA 4339

     114 SerAlaLeuPheSerAlaSerProMetAlaGlnProArgThrThrIleSe 130
         |||   ::::::::::::::::|||                   |||||
    4340 GAAGCCGAGCTTGATGGAGATCAAATG..................ATATC 4371

     130 rAspAlaGluIleTrpAspMetValSerGlnAsnIleSerAlaIleGlyA 147
         |:::   |||:::|||:::::::::::::::::|||:::::::|||:::
    4372 TCATAGAGAACTGTGGGCTAAAATTGCAAACTCCATCAATGATATTAATG 4421

     147 spSerTyrLeuGlyValTyrGluAsnValValAlaValTyrThrAspPhe 163
         ::   ||||||   ||||||||:::::::|||:::   ||||||:::
    4422 AACAGTATCTGAAAGTATATGAACATGCCGTTAGTTCATATACTCAAATG 4471

     164 TyrGlnAlaPheSerAspIleLeuSerLysMetGlyGlyTrpLeuLeuPr 180
         ||||||:::||||||:::::::||||||   ::::::||||||:::   ||
    4472 TATCAAGATTTTAGCGCTGTTCTTTCCAGTCTTGCCGGCTGGATCTCTCC 4521

     180 o...GlyLysAspGlyAsnThrValLysLeuAspValThrSerLeuLysA 196
         |   |||:::||||||||:::||||||||||:::|||:::||||||||:
    4522 CGGAGGTAACGACGGAAACTCCGTGAAATTACAAGTCAACTCGCTTAAAA 4571

     196 snAspLeuAsnSerLeuValAsnLysTyrAsnGlnIleAsnSerAsnThr 212
         :::::|||:::   |||   :::|||||:::             ::::::
    4572 AGGCATTGGAAGAACTCAAGGAAAAATATAAA............GATAAA 4609

     213 ValLeuPheProAlaGlnSerGlySerGlyValLysValAlaThrGluAl 229
         |||:::||||||:::              ::::::::::::::::::
    4610 CCGCTATATCCAGCAAAT..............AATACTGTTAGTCAGGA 4644

     229 aGluAlaArgGlnTrpLeuSerGluLeuAsnLeuProAsnSerCysLeuL 246
         :::|||   :::|||||:::||||||:::   :::   :::   :::
    4645 ACAAGCAAATAAATGGCTTACAGAATTAGGTGGAACAATCGGCAAGGTAT 4694

     246 ysSerTyrGlySerGlyTyrValValThrValAspLeuThrProLeuGln 262
              ::::::||||||||||||:::::::::::::||||||::::::
    4695 CTCAAAAAAACGGGGGATATGTTGTCAGTATAAACATGACCCCAATAGAC 4744

     263 LysMetValGlnAspIleAspGlyLeuGlyAlaProGlyLysAspSerLy 279
         :::|||:::::::::::::|||:::||||||:::          ::::::
    4745 AATATGTTAAAAAGCTTAGATAATCTAGGTGGA.........AATGGCGA 4785

     279 sLeuGluMetAspAsnAlaLysTyrGlnAlaTrpGlnSerGlyPheLysA 296
         :::   :::||||||||||||||||||||||||:::::::||||||   |
    4786 GGTTGTGCTAGATAATGCAAAATATCAGGCATGGAATGCCGGATTCTCTG 4835

     296 laGlnGluGluAsnMetLysThrThrLeuGlnThrLeuThrGlnLysTyr 312
         ||:::::::|||:::|||||:::::::|||||||||||:::||||||||
    4836 CCGAAGATGAAACAATGAAAAATAATCTTCAAACTTTAGTTCAAAAATAC 4885

     313 SerAsnAlaAsnSerLeuTyrAspAsnLeuValLysValLeuSerSerTh 329
         ||||||||||||||:::::::|||||||||||||||||||||||||||||
    4886 AGTAATGCCAATAGTATTTTTGATAATTTAGTAAAGGTTTTGAGTAGTAC 4935

     329 rIleSerSerSerLeuGluThrAlaLysSerPheLeuGln 342
         ||||||||||:::   :::|||:::|||   ||||||:::
    4936 AATAAGCTCATGTACAGATACAGATAAACTTTTTCTCCAT 4975

seq_name: gb_ba1:SDIPABCD

seq_documentation_block:
LOCUS       SDIPABCD     6869 bp    DNA             BCT       24-JUL-1991
DEFINITION  Shigella dysenteriae ipaBCD genes.
ACCESSION   X60777
NID         g46932
VERSION     X60777.1  GI:46932
KEYWORDS    ipaB gene; ipaC gene; ipaD gene; ippI gene.
SOURCE      Shigella dysenteriae.
  ORGANISM  Shigella dysenteriae
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Shigella.
REFERENCE   1  (bases 1 to 6869)
```

```
alignment_scores:
            Quality: 1237.00              Length:    344
              Ratio:  3.806                Gaps:      3
 Percent Similarity:  94.477   Percent Identity: 86.047

alignment_block:
 US-09-068-804-7 x STSIPGNS   ..

 Align seg 1/1  to: STSIPGNS  from: 1  to: 5393

       1 MetLeuAsnIleGlnAsnTyrSerAlaSerProHisProGlyIleValAl 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    3651 ATGCTTAATATTCAAAATTATTCCGCTTCTCCTCATCCGGGGATCGTTGC 3700

      17 aGluArgProGlnThrProSerAlaSerGluHisValGluThrAlaValV 34
         |||||||||||||||||||||||||||||||:::|||:::|||||||
    3701 CGAACGGCCGCAGACTCCTTCGGCGAGCGAGCACGCCGAGATTGCCGTGG 3750

      34 alProSerThrThrGluHisArgGlyThrAspIleIleSerLeuSerGln 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    3751 TACCGTCTACCACAGAACATCGCGGCACAGATATCATTTCATTATCGCAG 3800

      51 AlaAlaThrLysIleHisGlnAlaGlnGlnThrLeuGlnSerThrProPr 67
         ||||||||||||||:::|||||||||||||||||||||||||||||||||
    3801 GCGGCTACTAAAATCCAGCAGGCACAGCAGACGCTGCAGTCAACGCCACC 3850

      67 oIleSerGluGluAsnAsnAspGluArgThrLeuAlaArgGlnGlnLeuT 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    3851 GATTTCTGAAGAGAATAATGACGAGCGCACGCTGGCGCGCCAACAGTTGA 3900

      84 hrSerSerLeuAsnAlaLeuAlaLysSerGlyValSerLeuSerAlaGlu 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    3901 CCAGCAGCCTGAATGCGCTGGCGAAGTCCGGCGTGTCATTATCCGCAGAA 3950

     101 GlnAsnGluAsnLeuArgSerAlaPheSerAlaProThrSerAlaLeuPh 117
         ||||||||||||||||||:::|||||||||   ||||||||||||||||
    3951 CAAAATGAGAACCTGCGGAGCACGTTTTCTGCGCG.ACGTCGGCCTTATT 3999

     117 eSerAlaSerProMetAlaGlnPro.ArgThrThrIleSerAspAlaGlu 133
         ||||||||||||||||||   ::: ||||||||||||||||||||||||
    4000 TAGCGCTTCGCCTATGGCC...AGCGAGAACAACCATTTCTGATGCTGAG 4046

     134 IleTrpAspMetValSerGlnAsnIleSerAlaIleGlyAspSerTyrLe 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    4047 ATTTGGGATATGGTTTCCCAAAATATATCGGCGATAGGTGACAGCTACCT 4096

     150 uGlyValTyrGluAsnValValAlaValTyrThrAspPheTyrGlnAlaP 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    4097 GGGCGTTTATGAAAACGTTGTCGCAGTCTATACCGATTTTTATCAGGCCT 4146

     167 heSerAspIleLeuSerLysMetGlyGlyTrpLeuLeuProGlyLysAsp 183
         ||||||||||||||||||||||||||||||||   ||||||||||||
    4147 TCAGTGATATTCTTTCCAAAATGGGAGGCTGGTTATCGCCTGGTAAGGAT 4196

     184 GlyAsnThrValLysLeuAspValThrSerLeuLysAsnAspLeuAsnSe 200
         |||||||||:::||||||::|||   |||||||||:::::::::::||
    4197 GGAAATACCATTAAGCTAAATGTTGACTCACTTAAAAGTGAAATAAGTAG 4246

     200 rLeuValAsnLysTyrAsnGlnIleAsnSerAsnThrValLeuPheProA 217
         ||||:::|||||||||:::|||||||   ||||||:::|||||||||:
    4247 TTTAATTAATAAATACACTCAAATAAATAAAAATACGATTTTATTTCCCT 4296

     217 laGlnSerGlySerGlyValLysValAlaThrGluAlaGluAlaArgGln 233
         ::|||:::|||||||||:::::::::||||||   |||||||||   |||
    4297 CGCAAACTGGCAGCGGAATGACAACAGCAACGAAAGCGGAAGCTGAGCAG 4346

     234 TrpLeuSerGluLeuAsnLeuProAsnSerCysLeuLysSerTyrGlySe 250
         |||:::   |||||||||||||||:::||||||||||||:::   |||||
    4347 TGGATTAAAGAATTGAATTTACCGGACAGCTGTCTAAAGGCGTCTGGTTC 4396

     250 rGlyTyrValValThrValAspLeuThrProLeuGlnLysMetValGlnA 267
         ||||||||||||   ||||||      ||||||   |||||||||   |

    4397 TGGTTATGTCGTACTGGTGGATACGGGGCCACTGAGCAAAATGGTTAGCG 4446

     267 spIleAspGlyLeuGlyAlaProGlyLysAspSerLysLeuGluMetAsp 283
         ||:::::|||:::|||:::         :::|||   ||||||:::|||
    4447 ATCTTAATGGAATAGGATCG.........GGTTCAGCCCTTGAACTGGAT 4487

     284 AsnAlaLysTyrGlnAlaTrpGlnSerGlyPheLysAlaGlnGluGluAs 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    4488 AACGCCAAATATCAAGCCTGGCAGTCGGGTTTTAAAGCACAGGAAGAAAA 4537

     300 nMetLysThrThrLeuGlnThrLeuThrGlnLysTyrSerAsnAlaAsnS 317
         |:::||||||||||||||||||||||||||||||||||||||||||||||
    4538 TCTGAAAACCACATTACAGACGCTGACGCAAAAATATAGCAATGCCAATT 4587

     317 erLeuTyrAspAsnLeuValLysValLeuSerSerThrIleSerSerSer 333
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    4588 CATTGTACGACAACCTGGTAAAAGTGCTGAGCAGTACGATAAGTAGCAGC 4637

     334 LeuGluThrAlaLysSerPheLeuGlnGly 343
         |||||||||||||||||||||||||||||
    4638 CTGGAAACCGCCAAAAGCTTCCTGCAAGGA 4667

seq_name: gb_ba1:SHFIPABCD

seq_documentation_block:
LOCUS       SHFIPABCD    5255 bp    DNA             BCT       19-MAR-1991
DEFINITION  S.flexneri ipaBCD region encoding invasion plasmid antigens.
ACCESSION   J04117
NID         g152740
VERSION     J04117.1  GI:152740
KEYWORDS    ipaA gene; ipaB gene; ipaC gene; ipaD; membrane protein.
SOURCE      S.flexeri (strain M90T-W) DNA, clone pHC17.
  ORGANISM  Shigella flexneri
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Shigella.
REFERENCE   1  (bases 1 to 5255)
  AUTHORS   Venkatesan,M.M., Buysse,J.M. and Kopecko,D.J.
  TITLE     Characterization of invasion plasmid antigen genes (ipaBCD) from
            Shigella flexneri
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 85, 9317-9321 (1988)
  MEDLINE   89057927
COMMENT     Draft entry and clean copy of sequence [1] kindly submitted by
            M.M.Venkatesan, 03-JAN-1989.
FEATURES             Location/Qualifiers
     source          1. .5255
                     /organism="Shigella flexneri"
                     /db_xref="taxon:623"
     CDS             606. .1073
                     /note="18 kD protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA26521.1"
                     /db_xref="PID:g152741"
                     /db_xref="GI:152741"
                     /translation="MSLNITENESISTAVIDAINSGATLKDINAIPDDMMDDIYSYAY
                     DFYNKGRIEEAEVFFRFLCIYDFYNVDYIMGLAAIYQIKEQFQQAADLYAVAFALGKN
                     DYTPVFHTGQCQLRLKAPLKAKECFELVIQHSNDEKLKIKAQSYLDAIQDIKE"
     CDS             1079. .2821
                     /note="IpaB protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA26522.1"
                     /db_xref="PID:g152742"
                     /db_xref="GI:152742"
                     /translation="MHNVSTTTGFPLAKILTSTELGDNTIQAANDAANKLFSLTIAD
                     LTANQNINTTNAHSTSNILIPELKAPKSLNASSQLTILLIGNLIQILGEKSLTALTNKI
                     TAWKSQQQARQQKNLEFSDKINTLLSETEGLTRDYEKQINKLKNADSKIKDLENKINQ
                     IQTRLSNLDPESPEKKKLSREEIQLTIKKDAAVKDRTLIEQKTLSIHSKLTDKSMQLE
                     KEIDSFSAFSNTASAEQLSTQQKSLTGLASVTQLMATFIQLVGKNNEESLKNDLALFQ
                     SLQESRKTEMERKSDEYAAEVRKAEELNRVMGCVGKILGALLTIVSVVAAAFSGGASL
                     ALAAVGLALMVTDAIVQAATGNSFMEQALNPIMKAVIEPLIKLLSDAFTKMLEGLGVD
                     SKKAKMIGSILGAIAGALVLVAAVVLVATVGKQAAAKLAENIGKIIGKTLTDLIPKFL
                     KNFSSQLDDLITNAVARLNKFLGAAGDEVISKQIISTHLNQAVLLGESVNSATQAGGS
```

```
     234 rpLeuSerGluLeuAsnLeuProAsnSerCysLeuLysSerTyrGlySer 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     784 GGCTCAGTGAATTGAATTTACCGAATAGCTGCCTGAAATCTTATGGATCC 833

     251 GlyTyrValValThrValAspLeuThrProLeuGlnLysMetValGlnAs 267
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     834 GGTTATGTCGTCACCGTTGATCTGACGCCATTACAAAAAATGGTTCAGGA 883

     267 pIleAspGlyLeuGlyAlaProGlyLysAspSerLysLeuGluMetAspA 284
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     884 TATTGATGGTTTAGGCGCGCCGGGAAAAGACTCAAAACTCGAAATGGATA 933

     284 snAlaLysTyrGlnAlaTrpGlnSerGlyPheLysAlaGlnGluGluAsn 300
         ||||||||||||||||||||||:::||||||||||||||||||||||||
     934 ACGCCAAATATCAAGCCTGGCAGGCGGGTTTTAAAGCGCAGGAAGAAAAT 983

     301 MetLysThrThrLeuGlnThrLeuThrGlnLysTyrSerAsnAlaAsnSe 317
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     984 ATGAAAACCACATTACAGACGCTGACGCAAAAATATAGCAATGCCAATTC 1033

     317 rLeuTyrAspAsnLeuValLysValLeuSerSerThrIleSerSerSerL 334
         ||||||||||||||||||||||||||||||||||||||||||||||||   |
    1034 ATTGTACGACAACCTGGTAAAAGTGCTGAGCAGTACGATAAGTAGCAG.C 1082

     334 euGluThrAlaLysSerPheLeuGlnGly 343
         |||||||||||||||||||||||||||||
    1083 TGGAAACCGCCAAAAGCTTCCTGCAAGGA 1111

seq_name: gb_ba1:STSIPGNS

seq_documentation_block:
LOCUS       STSIPGNS     5393 bp    DNA             BCT       04-OCT-1995
DEFINITION  S.typhi spaT, sipB, sipC, sipD and sipA genes.
ACCESSION   X82670
NID         g1009001
VERSION     X82670.1  GI:1009001
KEYWORDS    sipA gene; sipB gene; sipC gene; sipD gene; spaT gene.
SOURCE      Salmonella typhi.
  ORGANISM  Salmonella typhi
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Salmonella.
REFERENCE   1  (bases 1 to 5393)
  AUTHORS   Hermant,D., Menard,R., Arricau,N., Parsot,C. and Popoff,M.Y.
  TITLE     Functional conservation of the Salmonella and Shigella effectors of
            entry into epithelial cells
  JOURNAL   Mol. Microbiol. 17 (4), 781-789 (1995)
  MEDLINE   96111497
REFERENCE   2  (bases 1 to 5393)
  AUTHORS   Popoff,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-1994) M. Popoff, Institut Pasteur, Unite des
            Enterobacteries, 28 rue du Docteur Roux, 75724 Paris Cedex 15,
            FRANCE
FEATURES             Location/Qualifiers
     source          1. .5393
                     /organism="Salmonella typhi"
                     /strain="Ty2"
                     /db_xref="taxon:601"
     gene            43. .540
                     /gene="spaT"
     CDS             43. .540
                     /gene="spaT"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA57987.1"
                     /db_xref="PID:g1009002"
                     /db_xref="GI:1009002"
                     /db_xref="SPTREMBL:Q57260"
                     /translation="MDYQNNVSEERVAEMIWDAVSEGATLKDVHGIPQDMMDGLYAHA
                     YEFYNQGRLDEAETFFRFLCIYDFYNPDYTMGLAAVCQLKKQFQKACDLYAVAFTLLK
                     NDYRPVFFTGQCQLLMRKAAKARQCFELVNERTEDESLRAKALVYLEALKTAETEQHS
                     EQEKE"
     gene            543. .2324
                     /gene="sipB"
     CDS             543. .2324
                     /gene="sipB"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA57988.1"
                     /db_xref="PID:g1009003"
                     /db_xref="GI:1009003"
                     /db_xref="SPTREMBL:Q56134"
                     /translation="MVNDASSISRSGYTQNPRLAEAAFEGVRKNTDFLKAADKAFKDV
                     VATKAGDLKAGTKSGESAINTVGLKPPTDAARKLSSEGQLTLLLGKLMTLLGDVSLS
                     QLESRLAVWQAMIESQKEMGIQVSKEFQTALGEAQEATDLYEASIKKTDTAKSVYDAA
                     AKKLTQAQNKLQSLDPADPGYAQAEAAVEQAGKEATEAKEALDKATDATVKAGTDAKA
                     KAEKADNILTKFQGTANAASQNQVSQGEQDNLSNVARLTMLMAMFIEIVGKNTEESLQ
                     NDLALFNALQEGRQAEMEKKSAEFQEETRKAEETNRIMGCIGKVLGALLTIVSVVAAV
                     FTGGASLALAAVGLAVMVADEIVKAATGVSFIQQALNPIMEHVLKPLMELIGKAITKA
                     LEGLGVDKKTAEMAGSIVGAIVAAIAMVAVIVVVAVVGKGAAAKLGNALSKMMGETIK
                     KLVPNVLKQLAQNGSKLFTQGMQRITSGLGNVGSKMGLQTNALSKELVGNTLNKVALG
                     MEVTNTAAQSAGGVAEGVFIKNASEALADFMLARFAMDQIQQWLKQSVEIFGENQKVT
                     AELQKAMSSAVQQNADASRFILRQSRA"
     gene            2352. .3581
                     /gene="sipC"
     CDS             2352. .3581
                     /gene="sipC"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA57989.1"
                     /db_xref="PID:g1009004"
                     /db_xref="GI:1009004"
                     /db_xref="SPTREMBL:Q56135"
                     /translation="MLISNVGINPAAYLNNHSVENSSQTASQSVSAKDILNSIGISSS
                     KVSDLGLSPTLSAPAPGVLTQTPGTITSFLKASIQNTDMNQDLNALANNVTTKANEVV
                     QTQLREQQAEVGKFFDISGMSSSAVALLAAANTLMLTLNQADSKLSGKLSLVSFDAAK
                     TTASSMMREGMNALSGSISQSALQLGITGVGAKLEYKGLQNERGALKHNAAKIDKLTT
                     ESHSIKNVLNGQNSVKLGAEGVDSLKSLNMKKTGTDATKNLNDATLKSNAGTSATESL
                     GIKNSNKQISPEHQAILSKRLESVESDIRLEQNTMDMTRIDARKMQMTGDLIMKNSVT
                     VGGIAGASRQYAATQERSEQQISQVNNRVASTASDEARESSRKSTSLIQEMLKTMESI
                     NQSKASALAAIAGNIRA"
     gene            3651. .4670
                     /gene="sipD"
     CDS             3651. .4670
                     /gene="sipD"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA57990.1"
                     /db_xref="PID:g1009005"
                     /db_xref="GI:1009005"
                     /db_xref="SPTREMBL:Q56136"
                     /translation="MLNIQNYSASPHPGIVAERPQTPSASEHAEIAVVPSTTEHRGTD
                     IISLSQAATKIQQAQQTLQSTPPISEENNDERTLARQQLTSSLNALAKSGVSLSAEQN
                     ENLRSTFSARRRPYLALRLWPARTTISDAEIWDMVSQNISAIGDSYLGVYENVVAVYT
                     DFYQAFSDILSKMGGWLSPGKDGNTIKLNVDSLKSEISSLINKYTQINKNTILFPSQT
                     GSGMTTATKAEAEQWIKELNLPDSCLKASGSGYVVLVDTGPLSKMVSDLNGIGSGSAL
                     ELDNAKYQAWQSGFKAQEENLKTTLQTLTQKYSNANSLYDNLVKVLSSTISSSLETAK
                     SFLQG"
     gene            4689. .5393
                     /gene="sipA"
     CDS             4689. .>5393
                     /gene="sipA"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA57991.1"
                     /db_xref="PID:g1009006"
                     /db_xref="GI:1009006"
                     /db_xref="SPTREMBL:Q56137"
                     /translation="MVTSVRTQPPVIMPGMQTEIKTQATNLAANLSAVRESATATLSG
                     EIKGQQLEDFPALIKQASLDALFKCGKDAEALKEVFTNSNNVAGKKAIMEFAGLFRSA
                     LNATSDSPEAKTLLMKVGAEYTAQIIKDGLKEKSAFGPWLPETKKAEAKLENLEKQLL
                     DIIKNNTGGELSKLSTNLVMQEVMPYIASCIEHNFGCTLDPLTRSSLTQLVDKAAAKA
                     VEALDMCHQKLTQEQGT"
BASE COUNT     1584 a    1156 c    1458 g    1195 t
ORIGIN
```

```
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Salmonella.
REFERENCE   1  (bases 1 to 3500)
  AUTHORS   Kaniga,K., Trollinger,D. and Galan,J.E.
  TITLE     Identification of two targets of the type III protein secretion
            system encoded by the inv and spa loci of Salmonella typhimurium
            that have homology to the Shigella IpaD and IpaA proteins
  JOURNAL   J. Bacteriol. 177 (24), 7078-7085 (1995)
  MEDLINE   96099289
REFERENCE   2  (bases 1 to 3500)
  AUTHORS   Galan,J.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-NOV-1995) Jorge E. Galan, Molecular Gentics and
            Microbiology, School of Medicine, SUNY Stony Brook, Stony Brook, NY
            11794-5222, USA
FEATURES             Location/Qualifiers
     source          1. .3500
                     /organism="Salmonella typhimurium"
                     /strain="SL1344"
                     /db_xref="taxon:602"
     gene            84. .1130
                     /gene="sipD"
     CDS             84. .1130
                     /gene="sipD"
                     /note="homolog of Shigella spp. IpaD"
                     /codon_start=1
                     /transl_table=11
                     /product="SipD"
                     /protein_id="AAA86617.1"
                     /db_xref="PID:g1172127"
                     /db_xref="GI:1172127"
                     /translation="MLNIQNYSASPHPGIVAERPQTPSASEHVETAVVPSTTEHRGTD
                     IISLSQAATKIHQAQQTLQSTPPISEENNDERTLARQQLTSSLNALAKSGVSLSAEQN
                     ENLRSAFSAPTSALFSASPMAQPRTTISDAEIWDMVSQNISAIGDSYLGVYENVVAVY
                     TDFYQAFSDILSKMGGWLLPGKDGNTVKLDVTSLKNDLNSLVNKYNQINSNTVLFPAQ
                     SGSGVKVATEAEARQWLSELNLPNSCLKSYGSGYVVTVDLTPLQKMVQDIDGLGAPGK
                     DSKLEMDNAKYQAWQAGFKAQEENMKTTLQTLTQKYSNANSLYDNLVKVLSSTISSSW
                     KPPKASCKDNRRGY"
     gene            1133. .3187
                     /gene="sipA"
     CDS             1133. .3187
                     /gene="sipA"
                     /note="homolog of Shigella spp. IpaA"
                     /codon_start=1
                     /transl_table=11
                     /product="SipA"
                     /protein_id="AAA86618.1"
                     /db_xref="PID:g1172128"
                     /db_xref="GI:1172128"
                     /translation="MVTSVRTQPPVIMPGMQTEIKTQATNLAANLSAVRAKCHSDAVR
                     EIKGPQLEDFPALIKQASLDALFKCGKDAEALKEVFTNSNNVAGKKAIMEFAGLFRSA
                     LNATSDSPEAKTLLMKVGAEYAAQIIKDGLKEKSAFGPWLPETKKAEAKLENLEKQLL
                     DIIKNNTGGELSKLSTNLVMQEVMPYIASCIEHNFGCTLDPLTRSNLTHLVDKAAAKA
                     VEALDMCPQKLTQEQGTSVGREARHLEMQTLIPLLLRNVFAQIPADKLPDPKIPEPAA
                     GPVPDGGKKAEPTGINININIDSSNHSVDNSKHINNAEPVDNGQRHIDNSNHDNSRKT
                     IDNSRTFIDNSQRNGESHHSTNSSNVSHSHSRVDSTTHQTETAHSASTGAIDHGIAGK
                     IDVTAHATAEAVTNASSESKDGKVVTSEKGTTGETTSFDEVDGVTSKSIIGKPVQATV
                     HGVDDNKQQSQTAEIVNVKPLASQLAGVENVKTDTLQSDTTVITGNKAGTTDNDNSQT
                     DKTGPFSGLKFKQNSFLSTVPSVTNMHSMHFDARETFLGVIRKALEPDTSTPFPVRRA
                     FDGLRAEILPNDTIKSAALKAQCSDIDKHPELKAKMETLKEVITHHPQKEKLAEIALQ
                     FAREAGLTRLKGETDYVLSNVLDGLIGDGSWRAGPAYESYLNKPGVDRVITTVDGLHM
                     QR"
     gene            3206. .3454
                     /gene="iacP"
     CDS             3206. .3454
                     /gene="iacP"
                     /note="IacP"
                     /codon_start=1
                     /transl_table=11
                     /product="acyl carrier protein"
                     /protein_id="AAA86619.1"
                     /db_xref="PID:g1172129"
                     /db_xref="GI:1172129"
                     /translation="MNMDIEARVKKVITSCIAVDVDSINGQTHLVEDLYADSLDLIDI
                     VFGLSEEFDISCNENDLPDMMTFADICRVVKKSLESRV"
BASE COUNT     1048 a    759 c    879 g    814 t
ORIGIN


alignment_scores:
            Quality: 1452.00              Length:    343
              Ratio:  4.233                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 99.708

alignment_block:
 US-09-068-804-7 x STU40013   ..

 Align seg 1/1  to: STU40013  from: 1  to: 3500

      1 MetLeuAsnIleGlnAsnTyrSerAlaSerProHisProGlyIleValAl 17
        |||||||||||||||||||||||||||||||||||||||||||||||||||
     84 ATGCTTAATATTCAAAATTATTCCGCTTCTCCTCATCCGGGGATCGTTGC 133

     17 aGluArgProGlnThrProSerAlaSerGluHisValGluThrAlaValV 34
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    134 CGAACGGCCGCAGACTCCCTCGGCGAGCGAGCACGTCGAGACTGCCGTGG 183

     34 alProSerThrThrGluHisArgGlyThrAspIleIleSerLeuSerGln 50
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    184 TACCGTCTACCACAGAACATCGCGGTACAGATATCATTTCATTATCGCAG 233

     51 AlaAlaThrLysIleHisGlnAlaGlnGlnThrLeuGlnSerThrProPr 67
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    234 GCGGCTACTAAAATCCACCAGGCACAGCAGACGCTGCAGTCAACGCCACC 283

     67 oIleSerGluGluAsnAsnAspGluArgThrLeuAlaArgGlnGlnLeuT 84
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    284 GATCTCTGAAGAGAATAATGACGAGCGCACGCTGGCGCGCCAGCAGTTGA 333

     84 hrSerSerLeuAsnAlaLeuAlaLysSerGlyValSerLeuSerAlaGlu 100
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    334 CCAGCAGCCTGAATGCGCTGGCGAAGTCCGGCGTGTCATTATCCGCAGAA 383

    101 GlnAsnGluAsnLeuArgSerAlaPheSerAlaProThrSerAlaLeuPh 117
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    384 CAAAATGAGAACCTGCGGAGCGCGTTTTCTGCGCCGACGTCGGCCTTATT 433

    117 eSerAlaSerProMetAlaGlnProArgThrThrIleSerAspAlaGluI 134
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    434 TAGCGCTTCGCCTATGGCGCAGCCGAGAACAACCATTTCTGATGCTGAGA 483

    134 leTrpAspMetValSerGlnAsnIleSerAlaIleGlyAspSerTyrLeu 150
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    484 TTTGGGATATGGTTTCCCAAAATATATCGGCGATAGGTGACAGCTATCTG 533

    151 GlyValTyrGluAsnValValAlaValTyrThrAspPheTyrGlnAlaPh 167
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    534 GGCGTTTATGAAAACGTTGTCGCAGTCTATACCGATTTTTATCAGGCCTT 583

    167 eSerAspIleLeuSerLysMetGlyGlyTrpLeuLeuProGlyLysAspG 184
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    584 CAGTGATATTCTTTCCAAAATGGGAGGCTGGTTATTACCAGGTAAGGACG 633

    184 lyAsnThrValLysLeuAspValThrSerLeuLysAsnAspLeuAsnSer 200
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    634 GTAATACCGTTAAGCTAGATGTTACCTCACTCAAAAATGATTTAAACAGT 683

    201 LeuValAsnLysTyrAsnGlnIleAsnSerAsnThrValLeuPheProAl 217
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    684 TTAGTCAATAAATATAATCAAATAAACAGTAATACCGTTTTATTTCCAGC 733

    217 aGlnSerGlySerGlyValLysValAlaThrGluAlaGluAlaArgGlnT 234
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    734 GCAGTCAGGCAGCGGCGTTAAAGTAGCCACTGAAGCGGAAGCGAGACAGT 783
```

```
                     GIKDSNKQISPEHQAILSKRLESVESDIRLEQNTMDMTRIDARKMQMTGDLIMKNSVT
                     VGGIAGASGQYAATQERSEQQISQVNNRVASTASDEARESSRKSTSLIQEMLKTMESI
                     NQSKASALAAIAGNIRA"
     gene            2198. .3229
                     /gene="sspD"
     CDS             2198. .3229
                     /gene="sspD"
                     /note="secreted protein"
                     /codon_start=1
                     /transl_table=11
                     /function="invasion determinant"
                     /evidence=experimental
                     /product="SspD"
                     /protein_id="AAC43548.1"
                     /db_xref="PID:g1117814"
                     /db_xref="GI:1117814"
                     /translation="MLNIQNYSASPHPGIVAERPQTPSASEHVETAVVPSTTEHRGTD
                     IISLSQAATKIHQAQQTLQSTPPISEENNDERTLARQQLTSSLNALAKSGVSLSAEQN
                     ENLRSAFSAPTSALFSASPMAQPRTTISDAEIWDMVSQNISAIGDSYLGVYENVVAVY
                     TDFYQAFSDILSKMGGWLLPGKDGNTVKLDVTSLKNDLNSLVNKYNQINSNTVLFPAQ
                     SGSGVKVATEAEARQWLSELNLPNSCLKSYGSGYVVTVDLTPLQKMVQDIDGLGAPGK
                     DSKLEMDNAKYQAWQSGFKAQEENMKTTLQTLTQKYSNANSLYDNLVKVLSSTISSSL
                     ETAKSFLQG"
     gene            3248. .3441
                     /gene="sspA"
     CDS             3248. .>3441
                     /gene="sspA"
                     /note="secreted protein"
                     /codon_start=1
                     /transl_table=11
                     /function="unknown"
                     /product="SspA"
                     /protein_id="AAC43549.1"
                     /db_xref="PID:g1117815"
                     /db_xref="GI:1117815"
                     /translation="MVTSVRTQPPVIMPGMQTEIKTQATNLAANLSAVRESATATLSG
                     EIKGPQLEDFPALIKQASLDA"
BASE COUNT      978 a    759 c    925 g    779 t
ORIGIN


alignment_scores:
           Quality: 1478.00            Length:    343
             Ratio:  4.309              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-068-804-7 x STU30491   ..

 Align seg 1/1  to: STU30491  from: 1  to: 3441

       1 MetLeuAsnIleGlnAsnTyrSerAlaSerProHisProGlyIleValAl 17
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    2198 ATGCTTAATATTCAAAATTATTCCGCTTCTCCTCATCCGGGGATCGTTGC 2247

      17 aGluArgProGlnThrProSerAlaSerGluHisValGluThrAlaValV 34
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    2248 CGAACGGCCGCAGACTCCCTCGGCGAGCGAGCACGTCGAGACTGCCGTGG 2297

      34 alProSerThrThrGluHisArgGlyThrAspIleIleSerLeuSerGln 50
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    2298 TACCGTCTACCACAGAACATCGCGGTACAGATATCATTTCATTATCGCAG 2347

      51 AlaAlaThrLysIleHisGlnAlaGlnGlnThrLeuGlnSerThrProPr 67
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    2348 GCGGCTACTAAAATCCACCAGGCACAGCAGACGCTGCAGTCAACGCCACC 2397

      67 oIleSerGluGluAsnAsnAspGluArgThrLeuAlaArgGlnGlnLeuT 84
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    2398 GATCTCTGAAGAGAATAATGACGAGCGCACGCTGGCGCGCCAGCAGTTGA 2447

      84 hrSerSerLeuAsnAlaLeuAlaLysSerGlyValSerLeuSerAlaGlu 100
         |||||||||||||||||||||||||||||||||||||||||||||||||||

    2448 CCAGCAGCCTGAATGCGCTGGCGAAGTCCGGCGTGTCATTATCCGCAGAA 2497

     101 GlnAsnGluAsnLeuArgSerAlaPheSerAlaProThrSerAlaLeuPh 117
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    2498 CAAAATGAGAACCTGCGGAGCGCGTTTTCTGCGCCGACGTCGGCCTTATT 2547

     117 eSerAlaSerProMetAlaGlnProArgThrThrIleSerAspAlaGluI 134
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    2548 TAGCGCTTCGCCTATGGCGCAGCCGAGAACAACCATTTCTGATGCTGAGA 2597

     134 leTrpAspMetValSerGlnAsnIleSerAlaIleGlyAspSerTyrLeu 150
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    2598 TTTGGGATATGGTTTCCCAAAATATATCGGCGATAGGTGACAGCTATCTG 2647

     151 GlyValTyrGluAsnValValAlaValTyrThrAspPheTyrGlnAlaPh 167
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    2648 GGCGTTTATGAAAACGTTGTCGCAGTCTATACCGATTTTTATCAGGCCTT 2697

     167 eSerAspIleLeuSerLysMetGlyGlyTrpLeuLeuProGlyLysAspG 184
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    2698 CAGTGATATTCTTTCCAAAATGGGAGGCTGGTTATTACCAGGTAAGGACG 2747

     184 lyAsnThrValLysLeuAspValThrSerLeuLysAsnAspLeuAsnSer 200
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    2748 GTAATACCGTTAAGCTAGATGTTACCTCACTCAAAAATGATTTAAACAGT 2797

     201 LeuValAsnLysTyrAsnGlnIleAsnSerAsnThrValLeuPheProAl 217
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    2798 TTAGTCAATAAATATAATCAAATAAACAGTAATACCGTTTTATTTCCAGC 2847

     217 aGlnSerGlySerGlyValLysValAlaThrGluAlaGluAlaArgGlnT 234
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    2848 GCAGTCAGGCAGCGGCGTTAAAGTAGCCACTGAAGCGGAAGCGAGACAGT 2897

     234 rpLeuSerGluLeuAsnLeuProAsnSerCysLeuLysSerTyrGlySer 250
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    2898 GGCTCAGTGAATTGAATTTACCGAATAGCTGCCTGAAATCTTATGGATCC 2947

     251 GlyTyrValValThrValAspLeuThrProLeuGlnLysMetValGlnAs 267
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    2948 GGTTATGTCGTCACCGTTGATCTGACGCCATTACAAAAAATGGTTCAGGA 2997

     267 pIleAspGlyLeuGlyAlaProGlyLysAspSerLysLeuGluMetAspA 284
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    2998 TATTGATGGTTTAGGCGCGCCGGGAAAAGACTCAAAACTCGAAATGGATA 3047

     284 snAlaLysTyrGlnAlaTrpGlnSerGlyPheLysAlaGlnGluGluAsn 300
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    3048 ACGCCAAATATCAAGCCTGGCAGTCGGGTTTTAAAGCGCAGGAAGAAAAT 3097

     301 MetLysThrThrLeuGlnThrLeuThrGlnLysTyrSerAsnAlaAsnSe 317
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    3098 ATGAAAACCACATTACAGACGCTGACGCAAAAATATAGCAATGCCAATTC 3147

     317 rLeuTyrAspAsnLeuValLysValLeuSerSerThrIleSerSerSerL 334
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    3148 ATTGTACGACAACCTGGTAAAAGTGCTGAGCAGTACGATAAGTAGCAGCC 3197

     334 euGluThrAlaLysSerPheLeuGlnGly 343
         |||||||||||||||||||||||||||||
    3198 TGGAAACCGCCAAAAGCTTCCTGCAAGGA 3226

seq_name: gb_bal:STU40013

seq_documentation_block:
LOCUS       STU40013     3500 bp    DNA             BCT       02-FEB-1996
DEFINITION  Salmonella typhimurium sipD, sipA, and acyl carrier protein (iacP)
            genes, complete cds.
ACCESSION   U40013
NID         g1172126
VERSION     U40013.1  GI:1172126
KEYWORDS    .
SOURCE      Salmonella typhimurium strain=SL1344.
```

OM of: US-09-068-804-7 to: GenEmbl:* out_format : pfs

Date: Oct 5, 1999 8:00 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09068804/runat_05101999_145134_10315/app_query.fasta.1
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=pam150 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=escore -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfs -NORM=stat -USER=US09068804 -NCPU=6 -ICPU=3 -WAIT
-THREADS=1

Search information block:
Query: US-09-068-804-7
Query length: 344
Database: GenEmbl:*
Database sequences: 679419
Database length: 1590154680
Search time (sec): 1424.920000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ..
gb_ba2:STU30491	+	1478.00	2667.12	1.3e-140	3441	! U30491 Salmonella typhimurium
gb_ba1:STU40013	+	1452.00	2619.63	5.7e-138	3500	! U40013 Salmonella typhimurium
gb_ba1:STSIPGNS	+	1237.00	2224.16	6.0e-116	5393	! X82670 S.typhi spaT, sipB, sip
gb_ba1:SHFIPABCD	+	508.00	897.09	5.0e-42	5255	! J04117 S.flexneri ipaBCD regic
gb_ba1:SDIPABCD	+	506.00	890.96	1.1e-41	6869	! X60777 Shigella dysenteriae ip
gb_ba1:SFPVR2	+	506.00	889.95	1.2e-41	7663	! X15319 Shigella flexineri plas
gb_ba1:W10PPIBCA	+	302.50	523.85	3.1e-21	4761	! M34849 Plasmid pWR100 (from S.
gb_ba1:ROPEPC	+	104.00	167.02	0.2314	2907	! X99379 R.obamensis PEPC gene.
gb_ba1:EFPPF54	+	89.00	143.65	4.63	1900	! X16421 E.faecium DNA encoding
gb_in1:DMPUMI	+	92.00	138.25	9.26	6120	! X62589 D.melanogaster mRNA for
gb_in1:DROPUM	+	92.00	137.23	10.55	6833	! L07943 Drosophila melanogaster
gb_pr2:HSCENPE	+	92.50	136.38	11.77	8257	! Z15005 H.sapiens CENP-E mRNA.
gb_pl1:GTU56117	+	80.00	136.00	12.36	742	! U56117 Ganoderma tsugae mangane
gb_ba1:BJA225023	+	90.50	135.92	12.49	5865	! AJ225023 Bradyrhizobium japoni
gb_in1:CEL131742	+	84.00	135.19	13.71	1773	! AJ131742 Caenorhabditis elegan
gb_ba2:AF139831	+	89.00	133.71	16.57	5542	! AF139831 Salmonella typhi prol
gb_ba1:KPU14004	+	82.50	132.88	18.44	1695	! U14004 Klebsiella pneumoniae C
gb_pl1:NCCOT1	+	83.50	131.85	21.06	2305	! X97657 N.crassa mRNA for serin
gb_in1:DDU86962	+	85.00	130.50	25.03	3576	! U86962 Dictyostelium discoideu
gb_ro:AB013130	+	82.50	130.44	25.21	2204	! AB013130 Rattus norvegicus mRN
gb_pl2:SCD9509	+	98.50	129.64	27.95	55359	! U32274 Saccharomyces cerevisi
gb_pr2:HSALL1X4	+	82.50	128.45	32.53	2730	! Z69747 H.sapiens ALL-1 gene ex
gb_pr1:AB018297	+	84.50	128.45	32.56	4044	! AB018297 Homo sapiens mRNA for
gb_pr1:HUMSP2A	+	81.00	128.32	33.08	2063	! M97190 Human Sp2 protein mRNA,
gb_vi:PPHBC	-	80.00	127.67	35.96	1819	! J02208 Human papillomavirus 1a
gb_ba1:MYCP115A	+	82.50	127.33	37.59	3082	! M34956 M.hyorhinis 115 kDa prc
gb_vi:SIVJ7P55	+	78.50	126.54	41.61	1532	! M74930 Simian immunodeficiency
gb_vi:HIVUG268	+	78.00	126.04	44.33	1465	! L11799 Human immunodeficiency
gb_ba2:RCDNAKJGE	+	82.50	126.01	44.50	3551	! U57637 Rhodobacter capsulatus
gb_pl1:ASNLP1A	+	79.50	125.96	44.77	1982	! L38317 Aspergillus niger srpA
gb_ba2:RCANIFF	-	74.00	124.65	53.01	777	! L42290 Rhodobacter capsulatus n
gb_ba2:AE001158	+	88.00	124.62	53.22	12134	! AE001158 Borrelia burgdorferi
gb_ba1:SOZ95912	+	71.00	124.57	53.53	435	! Z95912 Streptococcus oralis sod
gb_ba1:SOZ99194	+	71.00	124.57	53.53	435	! Z99194 Streptococcus oralis par
gb_ov:GGA16534	+	76.00	124.26	55.71	1199	! Y16534 Gallus gallus mRNA for
gb_pr1:HUMKG1D	+	81.00	124.00	57.63	3288	! D28588 Human mRNA for KIAA0048
gb_pr2:HSU91935	+	78.50	123.35	62.61	2159	! U91935 Human retina-derived PC
gb_ba2:AF082663	+	82.50	123.33	62.76	4740	! AF082663 Rhodococcus sp. NCIMB
gb_pl1:GTU56118	+	73.00	123.24	63.48	743	! U56118 Ganoderma tsugae mangane
gb_pl1:MICHSY	+	76.00	122.97	65.75	1378	! X17577 M.incana mRNA for chalc
gb_vi:SIVJ10P55	+	76.50	122.96	65.81	1521	! M74933 Simian immunodeficiency
gb_vi:AF110971	+	85.50	122.93	66.04	8908	! AF110971 HIV-1 isolate C-96BW1
gb_pr2:HSU91934	+	78.50	122.90	66.35	2267	! U91934 Human retina-derived
gb_vi:SIVJ11P55	+	76.50	122.89	66.42	1533	! M74934 Simian immunodeficie
gb_vi:SIVJ12P55	+	76.50	122.89	66.42	1533	! M74935 Simian immunodeficie

seq_name: gb_ba2:STU30491

seq_documentation_block:

```
LOCUS       STU30491     3441 bp    DNA              BCT       19-JAN-1996
DEFINITION  Salmonella typhimurium SspA (sspA) and SspB (sspB) genes, partial
            cds, and SspC (sspC) and SspD (sspD) genes, complete cds.
ACCESSION   U30491
NID         g1117811
VERSION     U30491.1  GI:1117811
KEYWORDS    .
SOURCE      Salmonella typhimurium.
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Salmonella.
REFERENCE   1  (bases 1 to 3441)
  AUTHORS   Hueck,C.J., Hantman,M.J., Bajaj,V., Johnston,C., Lee,C.A. and
            Miller,S.I.
  TITLE     Salmonella typhimurium secreted invasion determinants are
            homologous to Shigella Ipa proteins
  JOURNAL   Mol. Microbiol. 18 (3), 479-490 (1995)
  MEDLINE   96342378
REFERENCE   2  (bases 1 to 3441)
  AUTHORS   Miller,S.I.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUN-1995) Samuel I. Miller, Infectious Disease Unit,
            Massachusetts General Hospital, Harvard Medical School, Fruit
            Street, Boston, MA 02114, USA
FEATURES             Location/Qualifiers
     source          1. .3441
                     /organism="Salmonella typhimurium"
                     /strain="SL1344"
                     /db_xref="taxon:602"
                     /map="59 min"
     gene            1. .870
                     /gene="sspB"
     CDS             <1. .870
                     /gene="sspB"
                     /note="secreted protein"
                     /codon_start=1
                     /transl_table=11
                     /function="invasion determinant"
                     /product="SspB"
                     /protein_id="AAC43546.1"
                     /db_xref="PID:g1117812"
                     /db_xref="GI:1117812"
                     /translation="RKAEETNRIMGCIGKVLGALLTIVSVVAAVFTGGASLALAAVGL
                     AVMVADEIVKAATGVSFIQQALNPIMEHVLKPLMELIGKAITKALEGLGVDKKTAEMA
                     GSIVGAIVAAIAMVAVIVVVAVVGKGAAAKLGNALSKMMGETIKKLVPNVLKQLAQNG
                     SKLFTQGMQRITSGLGNVGSKMGLQTNALSKELVGNTLNKVALGMEVTNTAAQSAGGV
                     AEGVFIKNASEALADFMLARFAMDQIQQWLKQSVEIFGENQKVTAELQKAMSSAVQQN
                     ADASRFILRQSRA"
     gene            898. .2127
                     /gene="sspC"
     CDS             898. .2127
                     /gene="sspC"
                     /note="secreted protein"
                     /codon_start=1
                     /transl_table=11
                     /function="invasion determinant"
                     /evidence=experimental
                     /product="SspC"
                     /protein_id="AAC43547.1"
                     /db_xref="PID:g1117813"
                     /db_xref="GI:1117813"
                     /translation="MLISNVGINPAAYLNNHSVENSSQTASQSVSAKDILNSIGISSS
                     KVSDLGLSPTLSAPAPGVLTQTPGTITSFLKASIQNTDMNQDLNALANNVTTKANEVV
                     QTQLREQQAEVGKFFDISGMSSSAVALLAAANTLMLTLNQADSKLSGKLSLVSFDAAK
                     TTASSMMREGMNALSGSISQSALQLGITGVGAKLEYKGLQNERGALKHNAAKIDKLTT
                     ESHSIKNVLNGQNSVKLGAEGVDSLKSLNMKKTGTDATKNLNDATLKSNAGTSATESL
```

```
Db    181 GGGGTATTAACGCAAACCCCCGGAACGATCACGTCCTTTTTAAAAGCCAGTATTCAAAAT 240

Qy    241 ACCGACATGAATCAGGATTTGAATGCTCTGGCAAATAATGTCACGACTAAAGCGAATGAG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ACCGACATGAATCAGGATTTGAATGCTCTGGCAAATAATGTCACGACTAAAGCGAATGAG 300

Qy    301 GTTGTGCAAACCCAGTTACGCGAGCAGCAGGCAGAAGTCGGAAAGTTTTTTGATATTAGC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GTTGTGCAAACCCAGTTACGCGAGCAGCAGGCAGAAGTCGGAAAGTTTTTTGATATTAGC 360

Qy    361 GGAATGTCTTCCAGTGCCGTTGCGCTGTTGGCTGCCGCGAATACGTTAATGCTGACGTTG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GGAATGTCTTCCAGTGCCGTTGCGCTGTTGGCTGCCGCGAATACGTTAATGCTGACGTTG 420

Qy    421 AACCAGGCTGATAGCAAACTGTCTGGTAAGTTGTCATTAGTCAGTTTTGATGCAGCTAAA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AACCAGGCTGATAGCAAACTGTCTGGTAAGTTGTCATTAGTCAGTTTTGATGCAGCTAAA 480

Qy    481 ACGACGGCAAGCTCCATGATGCGCGAAGGGATGAATGCGTTGTCCGGTAGTATTTCCCAG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 ACGACGGCAAGCTCCATGATGCGCGAAGGGATGAATGCGTTGTCCGGTAGTATTTCCCAG 540

Qy    541 AGCGCGCTTCAGTTGGGGATCACTGGCGTGGGCGCCAAACTGGAATATAAGGGGCTGCAG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 AGCGCGCTTCAGTTGGGGATCACTGGCGTGGGCGCCAAACTGGAATATAAGGGGCTGCAG 600

Qy    601 AATGAAAGAGGCGCGCTTAAACATAATGCCGCGAAGATCGATAAACTGACCACTGAAAGC 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 AATGAAAGAGGCGCGCTTAAACATAATGCCGCGAAGATCGATAAACTGACCACTGAAAGC 660

Qy    661 CACAGTATTAAAAACGTGCTGAACGGGCAGAATAGCGTCAAACTCGGTGCTGAAGGCGTC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 CACAGTATTAAAAACGTGCTGAACGGGCAGAATAGCGTCAAACTCGGTGCTGAAGGCGTC 720

Qy    721 GATTCTCTGAAATCGTTAAATATGAAGAAAACCGGTACCGATGCGACGAAAAATCTTAAT 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 GATTCTCTGAAATCGTTAAATATGAAGAAAACCGGTACCGATGCGACGAAAAATCTTAAT 780

Qy    781 GATGCGACGCTTAAATCTAATGCCGGAACCAGCGCCACGGAAAGTCTGGGTATTAAAGAC 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GATGCGACGCTTAAATCTAATGCCGGAACCAGCGCCACGGAAAGTCTGGGTATTAAAGAC 840

Qy    841 AGTAATAAACAAATCTCCCCTGAACATCAGGCTATTCTGTCGAAACGTCTTGAGTCTGTC 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 AGTAATAAACAAATCTCCCCTGAACATCAGGCTATTCTGTCGAAACGTCTTGAGTCTGTC 900

Qy    901 GAATCCGATATTCGTCTTGAGCAGAATACCATGGATATGACCCGAATCGATGCGCGCAAG 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 GAATCCGATATTCGTCTTGAGCAGAATACCATGGATATGACCCGAATCGATGCGCGCAAG 960

Qy    961 ATGCAGATGACGGGCGATCTGATTATGAAGAACTCGGTCACGGTCGGTGGTATTGCAGGG 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 ATGCAGATGACGGGCGATCTGATTATGAAGAACTCGGTCACGGTCGGTGGTATTGCAGGG 1020

Qy   1021 GCGTCCGGGCAGTACGCCGCTACTCAGGAACGTTCCGAGCAGCAAATTAGCCAGGTGAAT 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 GCGTCCGGGCAGTACGCCGCTACTCAGGAACGTTCCGAGCAGCAAATTAGCCAGGTGAAT 1080

Qy   1081 AACCGGGTTGCCAGCACCGCATCGGACGAAGCCCGTGAAAGTTCACGTAAATCGACCAGC 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 AACCGGGTTGCCAGCACCGCATCGGACGAAGCCCGTGAAAGTTCACGTAAATCGACCAGC 1140

Qy   1141 CTGATTCAGGAAATGCTGAAAACAATGGAGAGCATTAACCAGTCGAAAGCATCCGCACTC 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 CTGATTCAGGAAATGCTGAAAACAATGGAGAGCATTAACCAGTCGAAAGCATCCGCACTC 1200

Qy   1201 GCTGCTATCGCAGGCAATATTCGCGCTTAA 1230
          |||||||||||||||||||||||||||||
Db   1201 GCTGCTATCGCAGGCAATATTCGCGCTTAA 1230
```

RESULT 2
T67035

```
ID   T67035 standard; DNA; 3441 BP.
AC   T67035;
DT   04-AUG-1997 (first entry)
DE   Salmonella virulence factor SspBCDA DNA region.
KW   Salmonella secreted protein; SspB; SspC; SspD; SspA;
KW   bacterial-mediated endocytosis; diagnosis; therapy; vaccine;
KW   attenuation; virulence; ss.
OS   Salmonella typhimurium.
FH   Key             Location/Qualifiers
FT   cds             1. .870
FT                   /*tag= a
FT                   /product= SspB (incomplete)
FT   cds             898. .2127
FT                   /*tag= b
FT                   /product= SspC
FT   cds             2198. .3228
FT                   /*tag= c
FT                   /product= SspD
FT   cds             3247. .3441
FT                   /*tag= d
FT                   /product= SspA (incomplete)
PN   WO9718225-A1.
PD   22-MAY-1997.
PF   14-NOV-1996; U18504.
PR   14-NOV-1995; US-006733.
PA   (GEHO ) GEN HOSPITAL CORP.
PI   Miller SI;
DR   WPI; 97-289217/26.
DR   P-PSDB; W15268-71.
PT   New isolated Salmonella secreted proteins and related genes - used
PT   to develop products for the detection, treatment or prevention of
PT   Salmonella infections
PS   Claim 11; Fig 28; 95pp; English.
CC   A DNA molecule (T67035) comprises the SspBCDA locus (truncated at
CC   the 5' and 3' ends) including the partial sspB, sspC, sspD and
CC   partial sspA sequences (see also T67028-31) coding for Salmonella
CC   typhimurium secreted proteins SspB, SspC, SspD and SspA (W15268-
CC   70). To identify the ssp genes, S. typhimurium mutants were
CC   investigated in invasion assays on HEp-2 epithelial cells. 2
CC   Ssp mutants with transposon insertions located between spaT and
CC   prgH (see also T67036) were identified. One mutant lacked Ssp of
CC   87, 42 and 36 kDa. The invasiveness of this mutant implicated the
CC   42 and 36 kDa Ssp in invasion. DNA sequencing of this region
CC   identified the 4 ssp (partial) genes. The ssp genes can be used in
CC   methods for diagnosing Salmonella infection and in methods of
CC   targetting an antigen to epithelial cells and of inducing a
CC   cytotoxic T cell immune response in a mammal. Ssp gene mutation
CC   can be used to create attenuated Salmonella strains useful as live
CC   vaccines.
SQ   Sequence 3441 BP; 978 A; 761 C; 925 G; 777 T;
```

Query Match 98.0%; Score 1205.4; DB 1; Length 3441;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1228; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

```
Qy      1 ATGTTAATTAGTAATGTGGGAATAAATCCCGCCGCTTATTTAAATAATCATTCTGTTGAG 60
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    898 ATGTTAATTACTAATGTGGGAATAAATCCCGCCGCTTATTTAAATAATCATTCTGTTGAG 957

Qy     61 AATAGTTCACAGACAGCTTCGCAATCCGTTAGCGCTAAAGATATTCTGAATAGTATTGGT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    958 AATAGTTCACAGACAGCTTCGCAATCCGTTAGCGCTAAAGATATTCTGAATAGTATTGGT 1017

Qy    121 ATTAGCAGCAGTAAAGTCAGTGACCTGGGGTTGAGTCCTACACTGAGCGCGCCTGCGCCA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1018 ATTAGCAGCAGTAAAGTCAGTGACCTGGGGTTGAGTCCTACACTGAGCGCGCCTGCGCCA 1077

Qy    181 GGGGTATTAACGCAAACCCCCGGAACGATCACGTCCTTTTTAAAAGCCAGTATTCAAAAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1078 GGGGTATTAACGCAAACCCCCGGAACGATCACGTCCTTTTTAAAAGCCAGTATTCAAAAT 1137

Qy    241 ACCGACATGAATCAGGATTTGAATGCTCTGGCAAATAATGTCACGACTAAAGCGAATGAG 300
```

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 1999, 17:58:06 ; Search time 223.92 Seconds
(without alignments)
1374.313 Million cell updates/sec

Title: US-09-068-804-2
Perfect score: 1230
Sequence: 1 ATGTTAATTAGTAATGTGGG.....CAGGCAATATTCGCGCTTAA 1230

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1230	100.0	1230	1	T67029	Salmonella secrete
2	1205.4	98.0	3441	1	T67035	Salmonella virulen
3	35	2.8	4940	1	Q91286	ORC1 gene encoding
4	35	2.8	4940	1	T73279	S. cerevisiae orig
5	34.4	2.8	7291	1	T41476	Haemophilus adhesi
6	34	2.8	459	1	Q63296	Consensus sequence
c 7	33.8	2.7	114955	1	X53491	Human adenosine A1
8	33.4	2.7	4709	1	V74898	Staphylococcus aur
9	33.2	2.7	1282	1	V73327	Human CD18 pseudog
10	32.8	2.7	513	1	Q23079	Antigen mc-30c gen
11	31.6	2.6	1612	1	N60392	Sequence encoding
12	31.6	2.6	565	1	Q35072	HCV envelope regio
13	31.6	2.6	1363	1	X14251	H. pylori GHPO 144
c 14	31.4	2.6	2973	1	Q66074	Sequence of micros
15	31.4	2.6	10592	1	T75616	Canine herpesvirus
16	31.4	2.6	10592	1	V66938	Canine herpes viru
c 17	31.4	2.6	10592	1	V66939	Canine herpes viru
c 18	31.4	2.6	4351	1	X16642	Coturnix sp. aroma
c 19	31	2.5	3172	1	X20524	Polynucleotide seq
20	30.6	2.5	940	1	N60906	Sequence encodes m
21	30.6	2.5	825	1	N60149	Sequence encoding
22	30.6	2.5	510	1	Q67870	Murine IL-2 DNA. A
23	30.6	2.5	1209	1	X34597	Murine IL-2/Ig fus
24	30.4	2.5	7900	1	V74449	Staphylococcus aur
25	30.2	2.5	237	1	V66943	Canine herpes viru
26	30	2.4	64976	1	V21209_16	Continuation (17 o
c 27	30	2.4	110000	1	X20248_00	Borrelia burgdorfe
c 28	29.8	2.4	896	1	T98648	DNA encoding a S.
c 29	29.8	2.4	757	1	X40005	Prostate cancer as
30	29.6	2.4	8247	1	N70996	Sequence encoding
c 31	29.6	2.4	110000	1	V21209_06	Continuation (7 of
32	29.6	2.4	13508	1	V74431	Staphylococcus aur
33	29.6	2.4	110000	1	X20248_06	Continuation (7 of
34	29.6	2.4	111309	1	X20250	Borrelia burgdorfe
35	29.6	2.4	1851	1	X19262	Staphylococcus aur
36	29.6	2.4	1440	1	X19263	Staphylococcus aur
37	29.4	2.4	2313	1	Q74708	cDNA of APP 770 ge
c 38	29.4	2.4	5350	1	T89254	cDNA for rat CASPR
39	29.4	2.4	510	1	V28775	Human interleukin-
c 40	29.4	2.4	1759	1	T84150	DNA encoding a pts
41	29.4	2.4	3494	1	V31270	E. coli J96 pathog
42	29.4	2.4	22671	1	V31200	E. coli J96 pathog
43	29.4	2.4	1521	1	V36456	Maltose binding pr
44	29.4	2.4	9183	1	V60751	HIV-1 strain YBF30
45	29.4	2.4	3045	1	V60754	HIV-1 strain YBF30

ALIGNMENTS

RESULT 1
T67029
ID T67029 standard; DNA; 1230 BP.
AC T67029;
DT 04-AUG-1997 (first entry)
DE Salmonella secreted protein SspC DNA.
KW Salmonella secreted protein; SspC; bacterial-mediated endocytosis;
KW diagnosis; therapy; vaccine; attenuation; virulence; ss.
OS Salmonella typhimurium.
FH Key Location/Qualifiers
FT cds 1. .1230
FT /*tag= a
FT /product= SspC
FT /note= "the SspC amino acid sequence deduced
FT from the coding sequence does not fully
FT correspond to the SspC sequence given in
FT Fig 13"
PN WO9718225-A1.
PD 22-MAY-1997.
PF 14-NOV-1996; U18504.
PR 14-NOV-1995; US-006733.
PA (GEHO) GEN HOSPITAL CORP.
PI Miller SI;
DR WPI; 97-289217/26.
DR P-PSDB; W15269.
PT New isolated Salmonella secreted proteins and related genes - used
PT to develop products for the detection, treatment or prevention of
PT Salmonella infections
PS Claim 5; Fig 20; 95pp; English.
CC A DNA molecule (T67029) codes for Salmonella secreted protein SspC
CC (W15269), a virulence factor involved in bacterial mediated
CC endocytosis, which is able to translocate into human epithelial
CC cells. To identify Ssp genes, S. typhimurium mutants were
CC investigated in invasion assays on HEp-2 epithelial cells. 2
CC Ssp mutants with transposon insertions located between spaT and
CC prgH were identified. Of these, one mutant lacked Ssp of 87,
CC 42 and 36 kDa. The invasiveness of this mutant implicated the
CC 42 and 36 kDa Ssp in invasion. DNA sequencing of this region (see
CC also T67035) identified 2 complete and 2 partial open reading
CC frames, designated sspB, sspC (T67029), sspD (T67030) and sspA
CC (T67031). These nucleic acids can be used in methods for diagnosing
CC Salmonella infection and in methods of targetting an antigen to
CC epithelial cells and of inducing a cytotoxic T cell immune response
CC in a mammal. Ssp gene mutation can be used to create attenuated
CC Salmonella strains useful as live vaccines.
SQ Sequence 1230 BP; 362 A; 270 C; 317 G; 281 T;

Query Match 100.0%; Score 1230; DB 1; Length 1230;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1230; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 ATGTTAATTAGTAATGTGGGAATAAATCCCGCCGCTTATTTAAATAATCATTCTGTTGAG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGTTAATTAGTAATGTGGGAATAAATCCCGCCGCTTATTTAAATAATCATTCTGTTGAG 60

Qy     61 AATAGTTCACAGACAGCTTCGCAATCCGTTAGCGCTAAAGATATTCTGAATAGTATTGGT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AATAGTTCACAGACAGCTTCGCAATCCGTTAGCGCTAAAGATATTCTGAATAGTATTGGT 120

Qy    121 ATTAGCAGCAGTAAAGTCAGTGACCTGGGGTTGAGTCCTACACTGAGCGCGCCTGCGCCA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ATTAGCAGCAGTAAAGTCAGTGACCTGGGGTTGAGTCCTACACTGAGCGCGCCTGCGCCA 180

Qy    181 GGGGTATTAACGCAAACCCCCGGAACGATCACGTCCTTTTTAAAAGCCAGTATTCAAAAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
US-08-946-617-2


  Query Match              2.4%;  Score 29;  DB 3;  Length 2448;
  Best Local Similarity   48.0%;  Pred. No. 10;
  Matches   83;  Conservative    0;  Mismatches   90;  Indels    0;  Gaps    0;

Qy     748 AAAACCGGTACCGATGCGACGAAAAATCTTAATGATGCGACGCTTAAATCTAATGCCGGA 807
           || ||   |    ||   |  |||| |   ||| |||  |   ||| || ||   | | |
Db    2001 AATACTTTTGATAATATCAAAAAAATTGAAAATCATGGTATATTTATATATATATCAGAA 2060

Qy     808 ACCAGCGCCACGGAAAGTCTGGGTATTAAAGACAGTAATAAACAAATCTCCCCTGAACAT 867
           | |  |  || ||   | | |    ||  ||    | |||| | || |  |||  |  |
Db    2061 GCTAAAGTAACTGATCCTTTTGAATATATCGATTCAATTAAATATATATTACCTACAGCT 2120

Qy     868 CAGGCTATTCTGTCGAAACGTCTTGAGTCTGTCGAATCCGATATTCGTCTTGA 920
           | |||| | |    |||| |  |     |    | ||   ||||  | | ||
Db    2121 AAAGCTAATTTTAATAAACCTTGTAGTATTTGTAATTCAACTATTGATGTAGA 2173


Search completed: October  5, 1999, 18:00:51
Job time: 11491 sec
```

```
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/659,984A
;      FILING DATE:  07-JUN-1996
;      CLASSIFICATION:  436
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/485,152
;      FILING DATE:  07-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Heslin, James M.
;      REGISTRATION NUMBER:  29,541
;      REFERENCE/DOCKET NUMBER:  15270-002810US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-326-2400
;      TELEFAX:  415-326-2422
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1521 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-659-984A-3


  Query Match             2.4%;  Score 29.4;  DB 4;  Length 1521;
  Best Local Similarity   52.0%;  Pred. No. 5.8;
  Matches   66;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy     193 CAAACCCCCGGAACGATCACGTCCTTTTTAAAAGCCAGTATTCAAAATACCGACATGAAT 252
           ||||  |  ||  ||  |  ||  |     |  | | || || ||| |||||||||  |
Db    1184 CAAATATCAAGACGGAGGAGATCTCTGAAGTGAATCTGGATGCAGAATTCCGACATGACT 1243

Qy     253 CAGGATTTGAATGCTCTGGCAAATAATGTCACGACTAAAGCGAATGAGGTTGTGCAAACC 312
           |||||| ||||     |    || |||       ||     |||   |   || |||||
Db    1244 CAGGATATGAAGTTCATCATCAAAAATTGGTGTTCTTTGCAGAAGATGTGGGTTCAAACA 1303

Qy     313 CAGTTAC 319
           || | |
Db    1304 AAGGTGC 1310


RESULT  14
US-08-526-964-2
; Sequence 2, Application US/08526964
; Patent No. 5698421
;  GENERAL INFORMATION:
;    APPLICANT:  Lambowitz, Alan M
;    APPLICANT:  Zimmerly, Steven
;    APPLICANT:  Guo, Huatao
;    APPLICANT:  Yang, Jian
;    TITLE OF INVENTION:  Nucleotide Integrase Preparation
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Calfee, Halter & Griswold
;      STREET:  800 Superior Avenue
;      CITY:  Cleveland
;      STATE:  Ohio
;      COUNTRY:  USA
;      ZIP:  44114
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/526,964
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Golrick, Mary E
;      REGISTRATION NUMBER:  34,829
;      REFERENCE/DOCKET NUMBER:  22727/00127
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (216) 622-8200
;      TELEFAX:  (216) 241-0816
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2448 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
US-08-526-964-2


  Query Match             2.4%;  Score 29;  DB 2;  Length 2448;
  Best Local Similarity   48.0%;  Pred. No. 10;
  Matches   83;  Conservative    0;  Mismatches   90;  Indels    0;  Gaps    0;

Qy     748 AAAACCGGTACCGATGCGACGAAAAATCTTAATGATGCGACGCTTAAATCTAATGCCGGA 807
           || ||   |   ||   |  |||| |   ||| ||| |   ||| || ||   | | |
Db    2001 AATACTTTTGATAATATCAAAAAAATTGAAAATCATGGTATATTTATATATATATCAGAA 2060

Qy     808 ACCAGCGCCACGGAAAGTCTGGGTATTAAAGACAGTAATAAACAAATCTCCCCTGAACAT 867
           | |  |  || ||   | | |    ||  ||    | |||| | || |  |||  |  |
Db    2061 GCTAAAGTAACTGATCCTTTTGAATATATCGATTCAATTAAATATATATTACCTACAGCT 2120

Qy     868 CAGGCTATTCTGTCGAAACGTCTTGAGTCTGTCGAATCCGATATTCGTCTTGA 920
           | |||| | |    |||| |  |      |    | ||   |||| | | ||
Db    2121 AAAGCTAATTTTAATAAACCTTGTAGTATTTGTAATTCAACTATTGATGTAGA 2173


RESULT  15
US-08-946-617-2
; Sequence 2, Application US/08946617
; Patent No. 5869634
;  GENERAL INFORMATION:
;    APPLICANT:  Lambowitz, Alan M
;    APPLICANT:  Zimmerly, Steven
;    APPLICANT:  Guo, Huatao
;    APPLICANT:  Yang, Jian
;    TITLE OF INVENTION:  Nucleotide Integrase Preparation
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Calfee, Halter & Griswold
;      STREET:  800 Superior Avenue
;      CITY:  Cleveland
;      STATE:  Ohio
;      COUNTRY:  USA
;      ZIP:  44114
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/946,617
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Golrick, Mary E
;      REGISTRATION NUMBER:  34,829
;      REFERENCE/DOCKET NUMBER:  22727/00127
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (216) 622-8200
;      TELEFAX:  (216) 241-0816
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2448 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
```

```
Db   1633 GGCGCACAGTAGGAGGTGGTACACCAGCAGTAATAAAAATAACCCAGGAAACAAGAAGTA 1692

Qy    215 CCTTTTTAAAAGCCAGTATTCAAAATACCGACATGAATCAGGATTTGAATG 265
Db   1693 GCAGTATTATGGGATAAATTTAACAAAAATGCTGAAAAAAAGAGTTATTTG 1743


RESULT 11
US-08-680-726A-59
; Sequence 59, Application US/08680726A
; Patent No. 5804197
;  GENERAL INFORMATION:
;    APPLICANT:  Haanes, Elizabeth J.
;    APPLICANT:  Frank, Rexann S.
;    TITLE OF INVENTION:  RECOMBINANT CANINE HERPESVIRUSES
;    NUMBER OF SEQUENCES:  92
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sheridan Ross & McIntosh
;      STREET:  1700 Lincoln Street, Suite 3500
;      CITY:  Denver
;      STATE:  Colorado
;      COUNTRY:  U.S.A.
;      ZIP:  80203
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/680,726A
;      FILING DATE:  12-JUL-1996
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Connell, Gary J.
;      REGISTRATION NUMBER:  32,020
;      REFERENCE/DOCKET NUMBER:  2618-46-C1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (303) 863-9700
;      TELEFAX:  (303) 863-0223
;  INFORMATION FOR SEQ ID NO:  59:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  237 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..234
US-08-680-726A-59


  Query Match             2.5%;  Score 30.2;  DB 3;  Length 237;
  Best Local Similarity   53.9%;  Pred. No. 1.1;
  Matches   62;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy      20 GAATAAATCCCGCCGCTTATTTAAATAATCATTCTGTTGAGAATAGTTCACAGACAGCTT 79
Db       3 GAATAAATCTAAACTCTCATTTAAAGAAAAAAACGCTATATATGAATTTAAAAATATTTT 62

Qy      80 CGCAATCCGTTAGCGCTAAAGATATTCTGAATAGTATTGGTATTAGCAGCAGTAA 134
Db      63 ATCAAACACTTCATTGTCAACTTTTCCTGTATTATCGTTTAATGAGGAGCCAAAA 117


RESULT 12
US-08-480-498-3
; Sequence 3, Application US/08480498
; Patent No. 5744346
;  GENERAL INFORMATION:
;    APPLICANT:  Chrysler, Susanna M.S.
;    APPLICANT:  Sinha, Sukanto
;    APPLICANT:  Keim, Pamela S.
;    APPLICANT:  Anderson, John P.
;    TITLE OF INVENTION:  Beta-Secretase
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend Khourie and Crew
;      STREET:  One Market Plaza, Steuart Tower, Suite 2000
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,498
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Heslin, James M.
;      REGISTRATION NUMBER:  29,541
;      REFERENCE/DOCKET NUMBER:  015270-002200
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-326-2400
;      TELEFAX:  415-326-2422
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1521 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-480-498-3


  Query Match             2.4%;  Score 29.4;  DB 2;  Length 1521;
  Best Local Similarity   52.0%;  Pred. No. 5.8;
  Matches   66;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy     193 CAAACCCCCGGAACGATCACGTCCTTTTTAAAAGCCAGTATTCAAAATACCGACATGAAT 252
Db    1184 CAAATATCAAGACGGAGGAGATCTCTGAAGTGAATCTGGATGCAGAATTCCGACATGACT 1243

Qy     253 CAGGATTTGAATGCTCTGGCAAATAATGTCACGACTAAAGCGAATGAGGTTGTGCAAACC 312
Db    1244 CAGGATATGAAGTTCATCATCAAAAATTGGTGTTCTTTGCAGAAGATGTGGGTTCAAACA 1303

Qy     313 CAGTTAC 319
Db    1304 AAGGTGC 1310


RESULT 13
US-08-659-984A-3
; Sequence 3, Application US/08659984A
; Patent No. 5942400
;  GENERAL INFORMATION:
;    APPLICANT:  Anderson, John P.
;    APPLICANT:  Sinha, Sukanto
;    APPLICANT:  Jacobson-Croak, Kirsten L.
;    TITLE OF INVENTION:  Assays for Detecting Beta-Secretase
;    TITLE OF INVENTION:  Inhibition
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend and Crew LLP
;      STREET:  Two Embarcadero Ctr., 8th Floor
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
```

```
;      LENGTH:  510 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-184-009-150


  Query Match             2.5%;  Score 30.6;  DB 3;  Length 510;
  Best Local Similarity   51.1%;  Pred. No. 1.3;
  Matches   72;  Conservative    0;  Mismatches   69;  Indels    0;  Gaps    0;

Qy     866 ATCAGGCTATTCTGTCGAAACGTCTTGAGTCTGTCGAATCCGATATTCGTCTTGAGCAGA 925
           | ||||| |     | ||||  |||| |||   | |||   ||  || | | |  || |
Db     245 AGCAGGCCACAGAATTGAAAGATCTTCAGTGCCTAGAAGATGAACTTGGACCTCTGCGGC 304

Qy     926 ATACCATGGATATGACCCGAATCGATGCGCGCAAGATGCAGATGACGGGCGATCTGATTA 985
           ||    ||||| |||| | || | |       |  || |     | |   || | || |
Db     305 ATGTTCTGGATTTGACTCAAAGCAAAAGCTTTCAATTGGAAGATGCTGAGAATTTCATCA 364

Qy     986 TGAAGAACTCGGTCACGGTCG 1006
           || | |   || || || |
Db     365 GCAATATCAGAGTAACTGTTG 385


RESULT  9
US-08-458-356-150
; Sequence 150, Application US/08458356
; Patent No. 5942235
;  GENERAL INFORMATION:
;    APPLICANT:  Paoletti, Enzo
;    APPLICANT:  Tartaglia, James
;    APPLICANT:  Cox, William I.
;    TITLE OF INVENTION:  RECOMBINANT VIRUS IMMUNOTHERAPY
;    NUMBER OF SEQUENCES:  217
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Curtis, Morris & Safford
;      STREET:  530 Fifth Avenue
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/458,356
;      FILING DATE:  02-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/184,009
;      FILING DATE:  19-JAN-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Frommer, William S.
;      REGISTRATION NUMBER:  25,506
;      REFERENCE/DOCKET NUMBER:  454310-2530
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 840-3333
;      TELEFAX:  (212) 840-0712
;      TELEX:  425066CURTMS
;  INFORMATION FOR SEQ ID NO:  150:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  510 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-458-356-150


  Query Match             2.5%;  Score 30.6;  DB 4;  Length 510;
  Best Local Similarity   51.1%;  Pred. No. 1.3;
  Matches   72;  Conservative    0;  Mismatches   69;  Indels    0;  Gaps    0;

Qy     866 ATCAGGCTATTCTGTCGAAACGTCTTGAGTCTGTCGAATCCGATATTCGTCTTGAGCAGA 925
           | ||||| |     | ||||  |||| |||   | |||   ||  || | | |  || |
Db     245 AGCAGGCCACAGAATTGAAAGATCTTCAGTGCCTAGAAGATGAACTTGGACCTCTGCGGC 304

Qy     926 ATACCATGGATATGACCCGAATCGATGCGCGCAAGATGCAGATGACGGGCGATCTGATTA 985
           ||    ||||| |||| | || | |       |  || |     | |   || | || |
Db     305 ATGTTCTGGATTTGACTCAAAGCAAAAGCTTTCAATTGGAAGATGCTGAGAATTTCATCA 364

Qy     986 TGAAGAACTCGGTCACGGTCG 1006
           || | |   || || || |
Db     365 GCAATATCAGAGTAACTGTTG 385


RESULT  10
US-08-539-304A-5
; Sequence 5, Application US/08539304A
; Patent No. 5792933
;  GENERAL INFORMATION:
;    APPLICANT:  MA, DIN-POW
;    TITLE OF INVENTION:  FIBER-SPECIFIC PROTEIN EXPRESSION IN THE
;    TITLE OF INVENTION:  COTTON PLANT
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT
;      STREET:  1755 JEFFERSON DAVIS HWY. SUITE 400
;      CITY:  ARLINGTON
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/539,304A
;      FILING DATE:  04-OCT-1995
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  NORMAN, OBLON F
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  2343-037-27
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-413-3000
;      TELEFAX:  703-413-2220
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2646 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(741..1093, 1220..1226)
;    FEATURE:
;      NAME/KEY:  intron
;      LOCATION:  1094..1219
US-08-539-304A-5


  Query Match             2.5%;  Score 30.2;  DB 3;  Length 2646;
  Best Local Similarity   48.5%;  Pred. No. 4.2;
  Matches   83;  Conservative    0;  Mismatches   88;  Indels    0;  Gaps    0;

Qy      95 CTAAAGATATTCTGAATAGTATTGGTATTAGCAGCAGTAAAGTCAGTGACCTGGGGTTGA 154
           ||||  |||     ||| || ||  |||||| |     ||  |  |||  ||   |
Db    1573 CTAAGAATAAAAGAAATGGTGTTTATATTAGTATTTAGAATCTTGATGAGTTGCTATACC 1632

Qy     155 GTCCTACACTGAGCGCGCCTGCGCCAGGGGTATTAACGCAAACCCCCGGAACGATCACGT 214
```

```
;    NUMBER OF SEQUENCES:  92
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sheridan Ross & McIntosh
;      STREET:  1700 Lincoln Street, Suite 3500
;      CITY:  Denver
;      STATE:  Colorado
;      COUNTRY:  U.S.A.
;      ZIP:  80203
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/680,726A
;      FILING DATE:  12-JUL-1996
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Connell, Gary J.
;      REGISTRATION NUMBER:  32,020
;      REFERENCE/DOCKET NUMBER:  2618-46-C1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (303) 863-9700
;      TELEFAX:  (303) 863-0223
;  INFORMATION FOR SEQ ID NO:  51:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10592 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-680-726A-51


  Query Match              2.6%;  Score 31.4;  DB 3;  Length 10592;
  Best Local Similarity   52.7%;  Pred. No. 3.6;
  Matches   68;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy      6 AATTAGTAATGTGGGAATAAATCCCGCCGCTTATTTAAATAATCATTCTGTTGAGAATAG 65
          ||| | |   |  |||||||||     | || |||||||  ||  |  |  |  | |  |
Db   9016 AATCAATCCAGAATGAATAAATCTAAACTCTCATTTAAAGAAAAAAACGCTATATATGAA 9075

Qy     66 TTCACAGACAGCTTCGCAATCCGTTAGCGCTAAAGATATTCTGAATAGTATTGGTATTAG 125
          || | | | |  ||  ||| |  ||      | ||  | | ||| ||  |  |   || ||
Db   9076 TTTAAAAATATTTTATCAAACACTTCATTGTCAACTTTTCCTGTATTATCGTTTAATGAG 9135

Qy    126 CAGCAGTAA 134
            |||   ||
Db   9136 GAGCCAAAA 9144


RESULT   7
US-08-680-726A-52/c
; Sequence 52, Application US/08680726A
; Patent No. 5804197
;  GENERAL INFORMATION:
;    APPLICANT:  Haanes, Elizabeth J.
;    APPLICANT:  Frank, Rexann S.
;    TITLE OF INVENTION:  RECOMBINANT CANINE HERPESVIRUSES
;    NUMBER OF SEQUENCES:  92
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sheridan Ross & McIntosh
;      STREET:  1700 Lincoln Street, Suite 3500
;      CITY:  Denver
;      STATE:  Colorado
;      COUNTRY:  U.S.A.
;      ZIP:  80203
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/680,726A
;      FILING DATE:  12-JUL-1996
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Connell, Gary J.
;      REGISTRATION NUMBER:  32,020
;      REFERENCE/DOCKET NUMBER:  2618-46-C1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (303) 863-9700
;      TELEFAX:  (303) 863-0223
;  INFORMATION FOR SEQ ID NO:  52:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10592 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-680-726A-52


  Query Match              2.6%;  Score 31.4;  DB 3;  Length 10592;
  Best Local Similarity   52.7%;  Pred. No. 3.6;
  Matches   68;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy      6 AATTAGTAATGTGGGAATAAATCCCGCCGCTTATTTAAATAATCATTCTGTTGAGAATAG 65
          ||| | |   |  |||||||||     | || |||||||  ||  |  |  |  | |  |
Db   1577 AATCAATCCAGAATGAATAAATCTAAACTCTCATTTAAAGAAAAAAACGCTATATATGAA 1518

Qy     66 TTCACAGACAGCTTCGCAATCCGTTAGCGCTAAAGATATTCTGAATAGTATTGGTATTAG 125
          || | | | |  ||  ||| |  ||      | ||  | | ||| ||  |  |   || ||
Db   1517 TTTAAAAATATTTTATCAAACACTTCATTGTCAACTTTTCCTGTATTATCGTTTAATGAG 1458

Qy    126 CAGCAGTAA 134
            |||   ||
Db   1457 GAGCCAAAA 1449


RESULT   8
US-08-184-009-150
; Sequence 150, Application US/08184009
; Patent No. 5833975
;  GENERAL INFORMATION:
;    APPLICANT:  Paoletti, Enzo
;    APPLICANT:  Tartaglia, James
;    APPLICANT:  Cox, William I.
;    TITLE OF INVENTION:  RECOMBINANT VIRUS IMMUNOTHERAPY
;    NUMBER OF SEQUENCES:  217
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Curtis, Morris & Safford
;      STREET:  530 Fifth Avenue
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/184,009
;      FILING DATE:  19-JAN-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Frommer, William S.
;      REGISTRATION NUMBER:  25,506
;      REFERENCE/DOCKET NUMBER:  454310-2530
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 840-3333
;      TELEFAX:  (212) 840-0712
;      TELEX:  425066CURTMS
;  INFORMATION FOR SEQ ID NO:  150:
;    SEQUENCE CHARACTERISTICS:
```

```
Qy     739 AATATGAAGAAAACCGGTACCGATGCGACGAAAAATCTTAATGATGCGACGCTTAAATCT 798
           ||   ||  ||||  ||| |  |   ||   | ||   || | |  || | |  | ||
Db    1334 AAAGAGATCAAAAGAGGTCCTCAAAAGAAAGATAAGGCTACTCAAACGGCACAAATTTCA 1393

Qy     799 AATGCCGGAACCAGCGCCACGGAAAGTCTGGGTATTAAAGACAGTAATAAACAAATCTCC 858
           | || | ||| || || || || |    || || | | ||| |||||| || |    ||
Db    1394 GACGCAGAAACAAGAGCTACAGATATAACGGATAATGAGGACGGTAATGAAGATGAATCA 1453

Qy     859 CCTGAACATCA 869
           ||||  || |
Db    1454 TCTGATTATGA 1464


RESULT   4
US-08-409-995-3
; Sequence 3, Application US/08409995
; Patent No. 5646259
;  GENERAL INFORMATION:
;    APPLICANT:  Barenkamp, Stephen I.
;    APPLICANT:  St. Geme III, Joseph W.
;    TITLE OF INVENTION:  Haemophilus Adhesion Proteins
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  Four Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/409,995
;      FILING DATE:  24-MAR-1995
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Silva, Robin M.
;      REGISTRATION NUMBER:  38,304
;      REFERENCE/DOCKET NUMBER:  A-61053/RFT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5738 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
US-08-409-995-3


  Query Match             2.8%;  Score 34.4;  DB 1;  Length 5738;
  Best Local Similarity   47.3%;  Pred. No. 0.26;
  Matches  104;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy     619 AAACATAATGCCGCGAAGATCGATAAACTGACCACTGAAAGCCACAGTATTAAAAACGTG 678
           ||||||| ||   |||     | |  |    |||  |   |  | || || |||| |
Db    3805 AAACATACTGTAACGATTGATGTTGCAGAAGCCAAAGTTGGTGATGGTCTTGAAAAAGAT 3864

Qy     679 CTGAACGGGCAGAATAGCGTCAAACTCGGTGCTGAAGGCGTCGATTCTCTGAAATCGTTA 738
           |||| ||| ||   ||||| | | |  |  || |  ||  |||   | |  |
Db    3865 ACTGACGGCAAGATTAAACTCAAAGTAGATAATACAGATGGGAATAATCTATTAACCGTT 3924

Qy     739 AATATGAAGAAAACCGGTACCGATGCGACGAAAAATCTTAATGATGCGACGCTTAAATCT 798
           ||   |  |||   |  ||| ||| | |    |  |||||  |  ||     |  |
Db    3925 GATGCAACAAAAGGTGCATCCGTTGCCAAGGGCGAGTTTAATGCCGTAACAACAGATGCA 3984

Qy     799 AATGCCGGAACCAGCGCCACGGAAAGTCTGGGTATTAAAG 838
           | | | |      || | |  | | |  | |   |||||
Db    3985 ACTACAGCCCAAGGCACAAATGCCAATGAGCGCGGTAAAG 4024


RESULT   5
US-08-387-942C-9
; Sequence 9, Application US/08387942C
; Patent No. 5939289
;  GENERAL INFORMATION:
;    APPLICANT:  ERTESVAG, HELGA
;    APPLICANT:  VALLA, SVEIN
;    APPLICANT:  SKJAK-BRAEK, GUDMUND
;    APPLICANT:  LARSEN, BJORN
;    TITLE OF INVENTION:  DNA COMPOUNDS COMPRISING SEQUENCES
;    TITLE OF INVENTION:  ENCODING MANNURONAN C-5-EPIMERASE
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BIRCH, STEWART, KOLASCH & BIRCH, LLP
;      STREET:  P.O.BOX 747
;      CITY:  FALLS CHURCH
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22042
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/387,942C
;      FILING DATE:  09-MAY-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MURPHY JR, GERALD M.
;      REGISTRATION NUMBER:  28,977
;      REFERENCE/DOCKET NUMBER:  1809-106P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-205-8000
;      TELEFAX:  703-205-8050
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  459 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-387-942C-9


  Query Match             2.6%;  Score 32.4;  DB 4;  Length 459;
  Best Local Similarity   50.0%;  Pred. No. 0.31;
  Matches   81;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy     552 GTTGGGGATCACTGGCGTGGGCGCCAAACTGGAATATAAGGGGCTGCAGAATGAAAGAGG 611
           | ||||  |||| ||  |||||| |       |    |    ||||| |      | |
Db     273 GCTGGGCTTCACCGGGCTGGGCGACGGCTACAACGGCACCCTGCTGCTGGTGCTCAACGC 332

Qy     612 CGCGCTTAAACATAATGCCGCGAAGATCGATAAACTGACCACTGAAAGCCACAGTATTAA 671
           |||    |  |  |    |  ||||| | |  |   |  | | ||  |||| |  |  |
Db     333 CGCCGGCACCCGCACCTACCTGAAGAGCTACGAGGCGGACGCCGAGGGCCAGCGCTTCGA 392

Qy     672 AAACGTGCTGAACGGGCAGAATAGCGTCAAACTCGGTGCTGA 713
           | ||  ||| |||| |  | | || | | ||||| || ||
Db     393 GATCGCCCTGGACGGCAACTACACCGGCCAGCTCGGCGCCGA 434


RESULT   6
US-08-680-726A-51
; Sequence 51, Application US/08680726A
; Patent No. 5804197
;  GENERAL INFORMATION:
;    APPLICANT:  Haanes, Elizabeth J.
;    APPLICANT:  Frank, Rexann S.
;    TITLE OF INVENTION:  RECOMBINANT CANINE HERPESVIRUSES
```

```
Db   1374 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1315

Qy    599 AGAATGAAAGAGGCGCGCTTAAACATAATGCCGCGAAGATCGATAAACTGACCACTGAAA 658
          :::: :::::::: : :   ::: : :: :  : :::::  :: :::  ::  :  ::::
Db   1314 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1255

Qy    659 GCCACAGTATTAAAAACGTGCTGAACGGGCAGAATAGCGTCAAACTCGGTGCTGAAGGCG 718
          :  : :: :  ::::: : :  ::: ::: :::: :: :  :::   :: :  ::::: :
Db   1254 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1195

Qy    719 TCGATTCTCTGAAATCGTTAAATATGAAGAAAACCGGTACCGATGCGACGAAAAATCTTA 778
            ::      ::::  :  ::: : ::::::::  :: :  :: : :: ::::::    :
Db   1194 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1135

Qy    779 ATGATGCGACGCTTAAATCTAATGCCGGAACCAGCGCCACGGAAAGTCTGGGTATTAAAG 838
          : :: : :: :   :::   :: :  ::::  :: :  : ::::::   ::: :  ::::
Db   1134 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1075

Qy    839 ACAGTAATAAACAAATCTCCCCTGAACATCAGGCTA 874
          : :: :: |    || |||||  || |  ||| | |
Db   1074 RRRRRRRRATCGCAAGCTCCCTCGACCTGCAGCCAA 1039


RESULT   2
US-08-484-105-1
; Sequence 1, Application US/08484105
; Patent No. 5589341
;  GENERAL INFORMATION:
;    APPLICANT:  STILLMAN, Bruce
;    APPLICANT:  BELL, Stephen P
;    APPLICANT:  KOBAYASHI, Ryuji
;    APPLICANT:  RINE, Jasper
;    APPLICANT:  FOSS, Margit
;    APPLICANT:  McNALLY, Francis J
;    APPLICANT:  LAURENSON, Patricia
;    APPLICANT:  HERSKOWITZ, Ira
;    APPLICANT:  LI, Joachim J
;    APPLICANT:  GAVIN, Kimberly
;    TITLE OF INVENTION:  ORIGIN OF REPLICATION COMPLEX GENES
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/484,105
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman Ph.D., Richard Aron
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  A-59032/DJB/RAO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 494-8700
;      TELEFAX:  (415) 494-8771
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4940 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-484-105-1

  Query Match             2.8%;  Score 35;  DB 1;  Length 4940;
  Best Local Similarity   54.2%;  Pred. No. 0.15;
  Matches   71;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy    739 AATATGAAGAAAACCGGTACCGATGCGACGAAAAATCTTAATGATGCGACGCTTAAATCT 798
          ||   ||  |||| ||| |  |   ||   | ||   || | |  || | |  |  ||
Db   1334 AAAGAGATCAAAAGAGGTCCTCAAAAGAAAGATAAGGCTACTCAAACGGCACAAATTTCA 1393

Qy    799 AATGCCGGAACCAGCGCCACGGAAAGTCTGGGTATTAAAGACAGTAATAAACAAATCTCC 858
          | || | ||| || || || || |    || || | | ||| |||||  || |    ||
Db   1394 GACGCAGAAACAAGAGCTACAGATATAACGGATAATGAGGACGGTAATGAAGATGAATCA 1453

Qy    859 CCTGAACATCA 869
          ||||  || |
Db   1454 TCTGATTATGA 1464


RESULT   3
US-08-484-106-1
; Sequence 1, Application US/08484106
; Patent No. 5614618
;  GENERAL INFORMATION:
;    APPLICANT:  STILLMAN, Bruce
;    APPLICANT:  BELL, Stephen P
;    APPLICANT:  KOBAYASHI, Ryuji
;    APPLICANT:  RINE, Jasper
;    APPLICANT:  FOSS, Margit
;    APPLICANT:  McNALLY, Francis J
;    APPLICANT:  LAURENSON, Patricia
;    APPLICANT:  HERSKOWITZ, Ira
;    APPLICANT:  LI, Joachim J
;    APPLICANT:  GAVIN, Kimberly
;    TITLE OF INVENTION:  ORIGIN OF REPLICATION COMPLEX GENES
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/484,106
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman Ph.D., Richard Aron
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  A-59032/DJB/RAO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 494-8700
;      TELEFAX:  (415) 494-8771
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4940 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-484-106-1


  Query Match             2.8%;  Score 35;  DB 1;  Length 4940;
  Best Local Similarity   54.2%;  Pred. No. 0.15;
  Matches   71;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;
```

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 1999, 18:00:38 ; Search time 140.41 Seconds
(without alignments)
911.426 Million cell updates/sec

Title: US-09-068-804-2
Perfect score: 1230
Sequence: 1 ATGTTAATTAGTAATGTGGG.....CAGGCAATATTCGCGCTTAA 1230

Scoring table: IDENTITY_NUC

Searched: 192659 seqs, 52021692 residues

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/5C_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/5D_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	50.8	4.1	7218	2	US-08-232-463-14	Sequence 14, Appl
	2	35	2.8	4940	1	US-08-484-105-1	Sequence 1, Appli
	3	35	2.8	4940	1	US-08-484-106-1	Sequence 1, Appli
	4	34.4	2.8	5738	1	US-08-409-995-3	Sequence 3, Appli
	5	32.4	2.6	459	4	US-08-387-942C-9	Sequence 9, Appli
	6	31.4	2.6	10592	3	US-08-680-726A-51	Sequence 51, Appl
c	7	31.4	2.6	10592	3	US-08-680-726A-52	Sequence 52, Appl
	8	30.6	2.5	510	3	US-08-184-009-150	Sequence 150, App
	9	30.6	2.5	510	4	US-08-458-356-150	Sequence 150, App
	10	30.2	2.5	2646	3	US-08-539-304A-5	Sequence 5, Appli
	11	30.2	2.5	237	3	US-08-680-726A-59	Sequence 59, Appl
	12	29.4	2.4	1521	2	US-08-480-498-3	Sequence 3, Appli
	13	29.4	2.4	1521	4	US-08-659-984A-3	Sequence 3, Appli
	14	29	2.4	2448	2	US-08-526-964-2	Sequence 2, Appli
	15	29	2.4	2448	3	US-08-946-617-2	Sequence 2, Appli
c	16	28.8	2.3	632	1	US-08-111-080-29	Sequence 29, Appl
c	17	28.8	2.3	632	1	US-08-111-080-31	Sequence 31, Appl
c	18	28.8	2.3	632	2	US-08-211-980-29	Sequence 29, Appl
c	19	28.8	2.3	632	2	US-08-211-980-31	Sequence 31, Appl
c	20	28.8	2.3	632	5	PCT-US93-07967-29	Sequence 29, Appl
c	21	28.8	2.3	632	5	PCT-US93-07967-31	Sequence 31, Appl
	22	28.6	2.3	14602	2	US-08-597-236-1	Sequence 1, Appli
	23	28.6	2.3	14602	3	US-08-746-682A-1	Sequence 1, Appli
	24	28.6	2.3	2631	3	US-08-510-215A-1	Sequence 1, Appli
	25	28.6	2.3	1764	3	US-08-642-684-1	Sequence 1, Appli
	26	28.6	2.3	1770	3	US-08-642-684-3	Sequence 3, Appli
	27	28.4	2.3	2761	4	US-08-436-664-19	Sequence 19, Appl
	28	28.4	2.3	2631	4	US-08-436-664-21	Sequence 21, Appl
	29	28.4	2.3	1764	4	US-08-436-664-22	Sequence 22, Appl
	30	28.4	2.3	1767	4	US-08-436-664-24	Sequence 24, Appl
	31	28.4	2.3	1773	4	US-08-436-664-26	Sequence 26, Appl
	32	28.4	2.3	2631	4	US-08-436-664-31	Sequence 31, Appl
	33	28.4	2.3	2631	4	US-08-436-664-33	Sequence 33, Appl
	34	28.4	2.3	2761	5	PCT-US95-04080-19	Sequence 19, Appl
	35	28.4	2.3	2631	5	PCT-US95-04080-21	Sequence 21, Appl
	36	28.4	2.3	1764	5	PCT-US95-04080-22	Sequence 22, Appl
	37	28.4	2.3	1767	5	PCT-US95-04080-24	Sequence 24, Appl
	38	28.4	2.3	1773	5	PCT-US95-04080-26	Sequence 26, Appl
	39	28.4	2.3	2631	5	PCT-US95-04080-31	Sequence 31, Appl
	40	28.4	2.3	2631	5	PCT-US95-04080-33	Sequence 33, Appl
	41	28.4	2.3	14311	5	PCT-US96-06053-1	Sequence 1, Appli
c	42	28.4	2.3	14311	5	PCT-US96-06053-7	Sequence 7, Appli
c	43	28.2	2.3	4615	1	US-08-188-582-1	Sequence 1, Appli
c	44	28.2	2.3	4615	1	US-08-646-715-1	Sequence 1, Appli
	45	28.2	2.3	8591	1	US-08-462-859A-6	Sequence 6, Appli

ALIGNMENTS

```
RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
;   APPLICANT: DORNER, F.
;   APPLICANT: SCHEIFLINGER, F.
;   APPLICANT: FALKNER, F. G.
;   TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
;   NUMBER OF SEQUENCES: 52
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Foley & Lardner
;     STREET: 1800 Diagonal Road, Suite 500
;     CITY: Alexandria
;     STATE: VA
;     COUNTRY: USA
;     ZIP: 22313-0299
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/232,463
;     FILING DATE:
;     CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US/07/935,313
;     FILING DATE:
;     APPLICATION NUMBER: EP 91 114 300.6
;     FILING DATE: 26-AUG-1991
;   ATTORNEY/AGENT INFORMATION:
;     NAME: BENT, Stephen A.
;     REGISTRATION NUMBER: 29,768
;     REFERENCE/DOCKET NUMBER: 30472/114 IMMU
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (703)836-9300
;     TELEFAX: (703)683-4109
;     TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 7218 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   IMMEDIATE SOURCE:
;     CLONE: pTZgpt-F1s
US-08-232-463-14


  Query Match             4.1%;  Score 50.8;  DB 2;  Length 7218;
  Best Local Similarity   4.0%;  Pred. No. 1.1e-06;
  Matches   16;  Conservative  219;  Mismatches  161;  Indels    0;  Gaps    0;

Qy     479 AAACGACGGCAAGCTCCATGATGCGCGAAGGGATGAATGCGTTGTCCGGTAGTATTTCCC 538
           ::: :: :: :::    : :: : : ::::::: ::: : :  :   :: :: :
Db    1434 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1375

Qy     539 AGAGCGCGCTTCAGTTGGGGATCACTGGCGTGGGCGCCAAACTGGAATATAAGGGGCTGC 598
           :::: : :    ::  ::::::  :  :: : ::: :  :::  :::: : ::::::  :
```

```
Qy     633 GAAGATCGATAAACTGACCACTGAAAGCCACAG 665
           | |   ||    | |||| | ||  || |||||
Db      63 GCAATGCGCAGGAATGACAATTGCTAGACACAG 31


RESULT  15
AI488791
LOCUS       AI488791      655 bp    mRNA            EST       12-APR-1999
DEFINITION  EST247130 tomato ovary, TAMU Lycopersicon esculentum cDNA clone
            cLED18H12, mRNA sequence.
ACCESSION   AI488791
NID         g4384162
VERSION     AI488791.1  GI:4384162
KEYWORDS    EST.
SOURCE      tomato.
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
            Potatoe; Lycopersicon.
REFERENCE   1  (bases 1 to 655)
  AUTHORS   Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T.,
            Chandra,I., Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman,C.L.,
            Nierman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley,S.D.
            and Giovannoni,J.
  TITLE     Generation of ESTs from tomato carpel tissue
  JOURNAL   Unpublished (1999)
COMMENT     On May 18, 1998 this sequence version replaced gi:3138155.

            Contact: David Frisch
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 4366
            Fax: 864 656 4293
            Email: dfrisch@CLEMSON.EDU.
FEATURES             Location/Qualifiers
     source          1. .655
                     /organism="Lycopersicon esculentum"
                     /cultivar="TA496"
                     /db_xref="taxon:4081"
                     /map="20"
                     /clone="cLED18H12"
                     /clone_lib="tomato ovary, TAMU"
                     /tissue_type="carpel"
                     /dev_stage="5 days pre-anthesis to 5 days post-anthesis"
                     /lab_host="XL1-Blue MRF'"
                     /note="Vector: pBlueScript SK(-); Site_1: EcoR1; Site_2:
                     Xho1; cLED - Tomato Carpel EST Library. OligodT-primed and
                     directionally cloned  cDNA in vector Lamda ZAP II with 5'
                     and 3' ends located at the EcoRI and  XhoI sites,
                     respectively."
BASE COUNT      193 a    122 c    160 g    180 t
ORIGIN


  Query Match             2.7%;  Score 33;  DB 47;  Length 655;
  Best Local Similarity   51.0%;  Pred. No. 12;
  Matches   78;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy     610 GGCGCGCTTAAACATAATGCCGCGAAGATCGATAAACTGACCACTGAAAGCCACAGTATT 669
           || ||  ||||| ||  || |  ||||| | | |  ||||||| | | | |  || | ||
Db     254 GGTGCTGTTAAATATGCTGACCTGAAGAACAACAGGCTGACCAATTACACCTTCAATTTT 313

Qy     670 AAAAACGTGCTGAACGGGCAGAATAGCGTCAAACTCGGTGCTGAAGGCGTCGATTCTCTG 729
             |  |  |||| |  |   ||   |    |    |   |         |   || |||
Db     314 GATCAGATGCTTAGTGATAAGGGAAATACCGCTGTTTATTTGCTCTATGCACATGCTCGT 373

Qy     730 AAATCGTTAAATATGAAGAAAACCGGTACCGAT 762
           |  |  | ||  || | |||| | ||||  |||
Db     374 ATTTGTTCAATCATCAGGAAATCTGGTAAAGAT 406
```

Search completed: October 5, 1999, 17:13:51
Job time: 8674 sec

```
Qy    910 ATTCGTCTTGAGCAGAATACCATGGATATGACCCGAATCGATGCGCGCAAGATGCAGATG 969
          || |||   |||||    | |||||   ||     |||   | ||   |   | || ||
Db    356 ATCCGTGAGGAGCAGCGGGCGATGGACCAGAAGATCATCCTGGAGCTGGACCGGAAGGTG 415

Qy    970 ACGGGCGATCTGATTATGAAGAACTCGGTCACGGTCGGTGG 1010
          | | | | | ||  |    | |   ||    ||| | |||
Db    416 GCTGACCAGCAGAGCACACTGGAGAAGGCGGGGGGTGGCTGG 456


RESULT  13
AI269888
LOCUS       AI269888      530 bp    mRNA            EST       28-JAN-1999
DEFINITION  qk04c01.x1 NCI_CGAP_Kid3 Homo sapiens cDNA clone IMAGE:1867968 3'
            similar to TR:O35347 O35347 DGCR6 PROTEIN ;, mRNA sequence.
ACCESSION   AI269888
NID         g3889055
VERSION     AI269888.1  GI:3889055
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 530)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Aug 21, 1998 this sequence version replaced.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 1028   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 459.
FEATURES             Location/Qualifiers
     source          1. .530
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1867968"
                     /clone_lib="NCI_CGAP_Kid3"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer,
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not
                     I and Eco RI sites of the modified pT7T3 vector. mRNA
                     source: 2 pooled kidneys. Library went through one round
                     of normalization. Library constructed by Bento Soares and
                     M. Fatima Bonaldo. "
BASE COUNT      120 a    162 c    174 g     74 t
ORIGIN


  Query Match              2.7%;  Score 33;  DB 44;  Length 530;
  Best Local Similarity   50.3%;  Pred. No. 11;
  Matches   81;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy    850 CAAATCTCCCCTGAACATCAGGCTATTCTGTCGAAACGTCTTGAGTCTGTCGAATCCGAT 909
          || | |   ||||    ||||||   || |  |  |   || ||| | || |||  |
Db    276 CACAACCTGCCTGTGGTTCAGGCGGCTCAGCAGCGAGAACTAGAGGCCGTGGAACACCGG 335

Qy    910 ATTCGTCTTGAGCAGAATACCATGGATATGACCCGAATCGATGCGCGCAAGATGCAGATG 969
          || |||   |||||    | |||||   ||     |||   | ||   |   | || ||
Db    336 ATCCGTGAGGAGCAGCGGGCGATGGACCAGAAGATCATCCTGGAGCTGGACCGGAAGGTG 395

Qy    970 ACGGGCGATCTGATTATGAAGAACTCGGTCACGGTCGGTGG 1010
          | | | | | ||  |    | |   ||    ||| | |||
Db    396 GCTGACCAGCAGAGCACACTGGAGAAGGCGGGGGGTGGCTGG 436


RESULT  14
AI392307/c
LOCUS       AI392307      503 bp    mRNA            EST       03-FEB-1999
DEFINITION  NCSM2G11T3 Subtracted Mycelial Neurospora crassa cDNA clone SM2G11
            5' similar to pyruvate kinase, mRNA sequence.
ACCESSION   AI392307
NID         g4220114
VERSION     AI392307.1  GI:4220114
KEYWORDS    EST.
SOURCE      Neurospora crassa.
  ORGANISM  Neurospora crassa
            Eukaryota; Fungi; Ascomycota; Euascomycetes; Pyrenomycetes;
            Sordariales; Sordariaceae; Neurospora.
REFERENCE   1  (bases 1 to 503)
  AUTHORS   Nelson,M.A., Kang,S., Braun,E.L., Crawford,M.E., Dolan,P.L.,
            Leonard,P.M., Mitchell,J., Armijo,A.M., Bean,L., Blueyes,E.,
            Cushing,T., Errett,A., Fleharty,M., Gorman,M., Judson,K.,
            Miller,R., Ortega,J., Pavlova,I., Perea,J., Todisco,S.,
            Trujillo,R., Valentine,J., Wells,A., Werner-Washburne,M., Yazzie,S.
            and Natvig,D.O.
  TITLE     Expressed sequences from conidial, mycelial, and sexual stages of
            Neurospora crassa
  JOURNAL   Fungal Genet. Biol. 21, 348-363 (1997)
  MEDLINE   97435549
COMMENT     On Jan 14, 1998 this sequence version replaced gi:1797068.

            Contact: Natvig,D.O./Nelson,M.A.
            Department of Biology
            University of New Mexico
            Castetter Hall, Albuquerque, NM 87131, USA
            Tel: 505 277 3411
            Fax: 505 277 0304
            Email: ngp@biology.unm.edu.
FEATURES             Location/Qualifiers
     source          1. .503
                     /organism="Neurospora crassa"
                     /strain="74-OR23-IV A (FGSC 2489)"
                     /db_xref="taxon:5141"
                     /clone="SM2G11"
                     /clone_lib="Subtracted Mycelial"
                     /sex="Mating type A"
                     /tissue_type="Mycelium"
                     /dev_stage="Mycelium"
                     /lab_host="E. coli"
                     /note="Vector: pBlueScript SK (-); Site_1: EcoRI; Site_2:
                     XhoI; 2% sucrose for 24 hours. cDNA directionally cloned
                     into pBluescript SK(-) using the Uni-ZAP XR vector system
                     (Stratagene, La Jolla, CA). Previously identified highly
                     expressed clones were subtracted from this library."
BASE COUNT      108 a    176 c     96 g    123 t
ORIGIN


  Query Match              2.7%;  Score 33;  DB 45;  Length 503;
  Best Local Similarity   51.0%;  Pred. No. 11;
  Matches   78;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy    513 GAATGCGTTGTCCGGTAGTATTTCCCAGAGCGCGCTTCAGTTGGGGATCACTGGCGTGGG 572
          | | | ||| ||  |  |||  |   || |||||     |  | | |||  ||  ||  |
Db    183 GGAGGTGTTATCGAGGTGTAGATTGAAGCGCGCGGGGTTGACGCGAATCTTTGAGGTATG 124

Qy    573 CGCCAAACTGGAATATAAGGGGCTGCAGAATGAAAGAGGCGCGCTTAAACATAATGCCGC 632
            |  || ||| || ||     ||| |||  | |   |     | || ||| |      ||
Db    123 GGGAAATCTGCAAGATCGACAGCTTCAGCGTCAGTAAATAAAGTTTCAACCTGGGCAAGC 64
```

```
Db    119 GCTGACCAGCAGAGCACACTGGAGAAGGCGGGGGTGGCTGG 79

RESULT 11
AA463473
LOCUS       AA463473      520 bp    mRNA            EST       10-JUN-1997
DEFINITION  zx98e09.s1 Soares_NhHMPu_S1 Homo sapiens cDNA clone IMAGE:811816 3'
            similar to TR:E228435 E228435 DGCR6. ;, mRNA sequence.
ACCESSION   AA463473
NID         g2188357
VERSION     AA463473.1  GI:2188357
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 520)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J.,
            Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-Merck EST Project 1997
  JOURNAL   Unpublished (1997)
COMMENT     On May 18, 1995 this sequence version replaced gi:811199.

            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Possible reversed clone: similarity on wrong strand
            Possible reversed clone: polyT not found
            Seq primer: -41m13 fwd. ET from Amersham
            High quality sequence stop: 472.
FEATURES             Location/Qualifiers
     source          1. .520
                     /organism="Homo sapiens"
                     /db_xref="GDB:6042727"
                     /db_xref="taxon:9606"
                     /map="5"
                     /clone="IMAGE:811816"
                     /clone_lib="Soares_NhHMPu_S1"
                     /tissue_type="Pooled human melanocyte, fetal heart, and
                     pregnant uterus"
                     /lab_host="DH10B"
                     /note="Organ: mixed (see below); Vector: pT7T3D-Pac
                     (Pharmacia) with a modified polylinker; Site_1: Not I;
                     Site_2: Eco RI; Equal amounts of plasmid DNA from three
                     normalized libraries (melanocyte 2NbHM, pregnant uterus
                     NbHPU, and fetal heart NbHH19W) were mixed, and ss circles
                     were made in vitro. Following HAP purification, this DNA
                     was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from pools of
                     5,000 clones made from the same 3 libraries. The pools
                     consisted of I.M.A.G.E. clones 260232-265223,
                     340488-345479, and 484488-489479."
BASE COUNT      115 a    159 c    172 g     74 t
ORIGIN


  Query Match             2.7%;  Score 33;  DB 34;  Length 520;
  Best Local Similarity   50.3%;  Pred. No. 11;
  Matches   81;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy     850 CAAATCTCCCCTGAACATCAGGCTATTCTGTCGAAACGTCTTGAGTCTGTCGAATCCGAT 909
           || | |   ||||    ||||||   || |  |  |   || ||| | || |||  |
Db     314 CACAACCTGCCTGTGGTTCAGGCGGCTCAGCAGCGAGAACTAGAGGCCGTGGAACACCGG 373

Qy     910 ATTCGTCTTGAGCAGAATACCATGGATATGACCCGAATCGATGCGCGCAAGATGCAGATG 969
```

```
           || |||   ||||||    | |||||   ||     |||   | ||   |   | || ||
Db     374 ATCCGTGAGGAGCAGCGGGCGATGGACCAGAAGATCATCCTGGAGCTGGACCGGAAGTTG 433

Qy     970 ACGGGCGATCTGATTATGAAGAACTCGGTCACGGTCGGTGG 1010
           | | | | | ||  |    | |   ||    |||  | |||
Db     434 GCTGACCAGCAGAGCACACTGGAGAAGGCGGGGGTGGCTGG 474


RESULT 12
AI016764
LOCUS       AI016764      535 bp    mRNA            EST       27-AUG-1998
DEFINITION  ov27d06.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1638539
            3' similar to TR:O35347 O35347 DGCR6 PROTEIN ;, mRNA sequence.
ACCESSION   AI016764
NID         g3231100
VERSION     AI016764.1  GI:3231100
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 535)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On May 18, 1995 this sequence version replaced gi:811023.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 1114   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 473.
FEATURES             Location/Qualifiers
     source          1. .535
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1638539"
                     /clone_lib="Soares_testis_NHT"
                     /sex="male"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from mRNA obtained from Clontech
                     Laboratories, Inc., and primed with  a Not I - oligo(dT)
                     primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTT 3']."
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization to Cot5, and was
                     constructed  by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      118 a    167 c    173 g     77 t
ORIGIN


  Query Match             2.7%;  Score 33;  DB 41;  Length 535;
  Best Local Similarity   50.3%;  Pred. No. 11;
  Matches   81;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy     850 CAAATCTCCCCTGAACATCAGGCTATTCTGTCGAAACGTCTTGAGTCTGTCGAATCCGAT 909
           || | |   ||||    ||||||   || |  |  |   || ||| | || |||  |
Db     296 CACAACCTGCCTGTGGTTCAGGCGGCTCAGCAGCGAGAACTAGAGGCCGTGGAACACCGG 355
```

```
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 478)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1402278.

            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:610182
            Seq primer: -28m13 rev2 ET from Amersham.
FEATURES             Location/Qualifiers
     source          1. .478
                     /organism="Mus musculus"
                     /strain="C3H"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1120846"
                     /clone_lib="Barstead mouse myotubes MPLRB5"
                     /cell_line="C2C12"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     [AATTCGGATCCTTG], digested with Not I and cloned into the
                     Not I  and Eco RI sites of the modified pT7T3 vector.
                     Library constructed  by Bob Barstead.  The C2C12 cell line
                     (available from ATCC,  catalog # CRL-1772) differentiates
                     rapidly, forming contractile  myotubes and producing
                     characteristic muscle proteins.  "
BASE COUNT      148 a    105 c    145 g     80 t
ORIGIN


  Query Match             2.7%;  Score 33.4;  DB 36;  Length 478;
  Best Local Similarity   47.0%;  Pred. No. 8;
  Matches  103;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy     577 AAACTGGAATATAAGGGGCTGCAGAATGAAAGAGGCGCGCTTAAACATAATGCCGCGAAG 636
           | || |  | | |  | |  | ||||  | |     ||    |     || ||   | |
Db     102 AGACAGACAGAGAGAGAGAAGAAGAAGAAGATCCTGGCAGAGATGAGGAAGGCGCTGGCG 161

Qy     637 ATCGATAAACTGACCACTGAAAGCCACAGTATTAAAAACGTGCTGAACGGGCAGAATAGC 696
           |||||  | ||||     ||     |  ||     ||    |  | |    |||||| ||
Db     162 ATCGACCACCTGAATGAAGACCAACTGAGAGAGAAGGCCAAGGAGCTGTGGCAGAGTATT 221

Qy     697 GTCAAACTCGGTGCTGAAGGCGTCGATTCTCTGAAATCGTTAAATATGAAGAAAACCGGT 756
           |||  || |  |||||     ||||    | | ||  ||| ||   | |||||  ||
Db     222 CACAACCTGGAGGCTGAGAAGTTCGACCTGCAGGAAAAGTTCAAGCAGCAGAAATACGAA 281

Qy     757 ACCGATGCGACGAAAAATCTTAATGATGCGACGCTTAAA 795
           | | | |    | |||  |  |   |||  |  |  |||
Db     282 ATCAACGTTCTGCGAAACCGGATCAATGACAACCAGAAA 320


RESULT  10
AA234615/c
LOCUS       AA234615      439 bp    mRNA            EST       07-AUG-1997
DEFINITION  zs39d02.r1 Soares_NhHMPu_S1 Homo sapiens cDNA clone IMAGE:687555 5'
            similar to TR:E228435 E228435 DGCR6. ;, mRNA sequence.
ACCESSION   AA234615
NID         g1859109
VERSION     AA234615.1  GI:1859109
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 439)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,
            Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
            Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
            Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J.,
            Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.
            and Marra,M.
  TITLE     Generation and analysis of 280,000 human expressed sequence tags
  JOURNAL   Genome Res. 6 (9), 807-828 (1996)
  MEDLINE   97044478
COMMENT     On Dec 30, 1996 this sequence version replaced gi:1528985.

            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Possible reversed clone: similarity on wrong strand
            Insert Length: 591   Std Error: 0.00
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 387.
FEATURES             Location/Qualifiers
     source          1. .439
                     /organism="Homo sapiens"
                     /db_xref="GDB:5591705"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:687555"
                     /clone_lib="Soares_NhHMPu_S1"
                     /tissue_type="Pooled human melanocyte, fetal heart, and
                     pregnant uterus"
                     /lab_host="DH10B"
                     /note="Organ: mixed (see below); Vector: pT7T3D-Pac
                     (Pharmacia) with a modified polylinker; Site_1: Not I;
                     Site_2: Eco RI; Equal amounts of plasmid DNA from three
                     normalized libraries (melanocyte 2NbHM, pregnant uterus
                     NbHPU, and fetal heart NbHH19W) were mixed, and ss circles
                     were made in vitro. Following HAP purification, this DNA
                     was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from pools of
                     5,000 clones made from the same 3 libraries. The pools
                     consisted of I.M.A.G.E. clones 260232-265223,
                     340488-345479, and 484488-489479."
BASE COUNT       60 a    140 c    135 g    104 t
ORIGIN


  Query Match             2.7%;  Score 33;  DB 30;  Length 439;
  Best Local Similarity   50.3%;  Pred. No. 10;
  Matches   81;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy     850 CAAATCTCCCCTGAACATCAGGCTATTCTGTCGAAACGTCTTGAGTCTGTCGAATCCGAT 909
           || | |   ||||    ||||||   || |  |  |   || ||| | || |||  |
Db     239 CACAACCTGCCTGTGGTTCAGGCGGCTCAGCAGCGAGAACTAGAGGCCGTGGAACACCGG 180

Qy     910 ATTCGTCTTGAGCAGAATACCATGGATATGACCCGAATCGATGCGCGCAAGATGCAGATG 969
           || |||   ||||||    | |||||  ||     |||   | ||   |   | || ||
Db     179 ATCCGTGAGGAGCAGCGGGCGATGGACCAGAAGATCATCCTGGAGCTGGACCGGAAGGTG 120

Qy     970 ACGGGCGATCTGATTATGAAGAACTCGGTCACGGTCGGTGG 1010
```

```
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1309707.

            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:352777
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 406.
FEATURES             Location/Qualifiers
     source          1. .450
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /map="6 cen-q14"
                     /clone="IMAGE:578129"
                     /clone_lib="Soares 2NbMT"
                     /sex="male"
                     /tissue_type="Thymus"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGTTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. RNA
                     provided by Dr. Bertrand Jordan. Library went through two
                     rounds of normalization, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT       98 a    107 c    106 g    139 t
ORIGIN


  Query Match             2.7%;  Score 33.6;  DB 28;  Length 450;
  Best Local Similarity   50.0%;  Pred. No. 6.8;
  Matches   84;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy     875 TTCTGTCGAAACGTCTTGAGTCTGTCGAATCCGATATTCGTCTTGAGCAGAATACCATGG 934
           | ||| |    |    || |||||||   | | | ||||| |||| |  |||   ||  |
Db      66 TCCTGCCTTTGCTGTGTGCGTCTGTCAGCTGCAAGATTCGCCTTGTGGTGAAATGCACAG 125

Qy     935 ATATGACCCGAATCGATGCGCGCAAGATGCAGATGACGGGCGATCTGATTATGAAGAACT 994
           | | || |    ||    |  |  |  | |  |    | |  |  ||||  | ||||||
Db     126 AGCTCACACTGGGCGGAAGGGACGTGCGGAACTTAGATGTCTCTGGGATTCAGGAGAACT 185

Qy     995 CGGTCACGGTCGGTGGTATTGCAGGGGCGTCCGGGCAGTACGCCGCTA 1042
           |     |    | |   |  |||| ||| || | |      |||||||
Db     186 CTTATCCCTGAGATTAAAAAGCAGAGGCCTCAGAGGGAGCAGCCGCTA 233


RESULT   8
AI073883
LOCUS       AI073883      452 bp    mRNA            EST       13-FEB-1999
DEFINITION  oo12c04.x1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone
            IMAGE:1565958 3' similar to SW:GDL_DROME P22468 GONADAL PROTEIN
            GDL. ;, mRNA sequence.
ACCESSION   AI073883
NID         g3400527
VERSION     AI073883.1  GI:3400527
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 452)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 14, 1998 this sequence version replaced gi:1798811.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1128   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham.
FEATURES             Location/Qualifiers
     source          1. .452
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1565958"
                     /clone_lib="Soares_NSF_F8_9W_OT_PA_P_S1"
                     /lab_host="DH10B"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from five normalized
                     libraries were mixed, and ss circles were made in vitro.
                     Following HAP purification, this DNA was used as tracer in
                     a subtractive hybridization reaction. The driver was
                     PCR-amplified cDNAs from pools of 5,000 clones made from
                     the same 5 libraries. The pools consisted of the following
                     libraries and cloneIDs: Soares NbHSF pool 1:
                     309384-310919, 323208-325895 Soares Nb2HP pool 1:
                     145032-147335, 147720-148103, 148872-149255, 15002 -
                     150407, 151176-152327 Soares Nb2HF8-9W pool 1:
                     758280-760583, 772104-774407 Soares NbHPA pool 1:
                     304776-306311, 320136-322823, 326280-326663 Soares NbHOT
                     pool 1:  723720-726407, 739080-740999 Subtraction by Bento
                     Soares and M. Fatima Bonaldo."
BASE COUNT       99 a    142 c    147 g     63 t      1 others
ORIGIN


  Query Match             2.7%;  Score 33.6;  DB 42;  Length 452;
  Best Local Similarity   50.3%;  Pred. No. 6.8;
  Matches   81;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy     850 CAAATCTCCCCTGAACATCAGGCTATTCTGTCGAAACGTCTTGAGTCTGTCGAATCCGAT 909
           || | |   ||||   ||||||  || |  |  |   || ||| | || |||  |
Db     286 CACAACCTGCCTGTGGTTCAGGCGGCTCAGCAGCGAGAACTAGAGGCCGTGGAACACCGG 345

Qy     910 ATTCGTCTTGAGCAGAATACCATGGATATGACCCGAATCGATGCGCGCAAGATGCAGATG 969
           || |||   ||||||    | |||||   ||     |||   | ||   |   | || ||
Db     346 ATCCGTGAGGAGCAGCGGGCGATGGACCAGAAGATCATCCTGGAGCTGGACCGGAAGGTG 405

Qy     970 ACGGGCGATCTGATTATGAAGAACTCGGTCACGGTCGGTGG 1010
           | | | | | ||  |    | |   ||    ||| | |||
Db     406 GCTGACCAGCAGAGCACACTGGAGAAGGCGGNGGTGGCTGG 446


RESULT   9
AA636805
LOCUS       AA636805      478 bp    mRNA            EST       22-OCT-1997
DEFINITION  vr16e12.r1 Barstead mouse myotubes MPLRB5 Mus musculus cDNA clone
            IMAGE:1120846 5' similar to gb:M19309 TROPONIN T, SLOW SKELETAL
            MUSCLE ISOFORMS (HUMAN);, mRNA sequence.
ACCESSION   AA636805
NID         g2560584
VERSION     AA636805.1  GI:2560584
KEYWORDS    EST.
```

```
COMMENT     On Jan 17, 1998 this sequence version replaced gi:1900104.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 1074   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 438.
FEATURES             Location/Qualifiers
     source          1. .597
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1870612"
                     /clone_lib="NCI_CGAP_Kid3"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer,
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not
                     I and Eco RI sites  of the modified pT7T3 vector.  mRNA
                     source: 2 pooled kidneys. Library went through one round
                     of normalization. Library constructed by Bento Soares and
                     M. Fatima Bonaldo.  "
BASE COUNT      130 a    183 c    199 g     83 t      2 others
ORIGIN


  Query Match             2.8%;  Score 34.4;  DB 44;  Length 597;
  Best Local Similarity   45.2%;  Pred. No. 4.3;
  Matches  122;  Conservative    0;  Mismatches  148;  Indels    0;  Gaps    0;

Qy     850 CAAATCTCCCCTGAACATCAGGCTATTCTGTCGAAACGTCTTGAGTCTGTCGAATCCGAT 909
           || | |   ||||    ||||||   || |  |  |   || ||| | || |||  |
Db     316 CACAACCTGCCTGTGGTTCAGGCGGCTCAGCAGCGAGAACTAGAGGCCGTGGAACACCGG 375

Qy     910 ATTCGTCTTGAGCAGAATACCATGGATATGACCCGAATCGATGCGCGCAAGATGCAGATG 969
           || |||   ||||||    | |||||   ||     |||   | ||   |   | || ||
Db     376 ATCCGTGAGGAGCAGCGGGCGATGGACCAGAAGATCATCCTGGAGCTGGACCGGAAGGTG 435

Qy     970 ACGGGCGATCTGATTATGAAGAACTCGGTCACGGTCGGTGGTATTGCAGGGGCGTCCGGG 1029
           | | | | | ||  |    | |   | |  ||| | |||  |    | | |  ||
Db     436 GCTGACCAGCAGAGCACACTGGAGAAAGGCGGGGGTGGCTGGCTTCTACGTGACCACCAAC 495

Qy    1030 CAGTACGCCGCTACTCAGGAACGTTCCGAGCAGCAAATTAGCCAGGTGAATAACCGGGTT 1089
           |   | |     |  | | |       |  || | || | |  |    |    | | |
Db     496 CCACAGGAGCTGATGCTGCAGATGAACCTGCTGGAACTCATNCGAAAGCTGCAGCAGAGG 555

Qy    1090 GCCAGCACCGCATCGGACGAAGCCCGTGAA 1119
           | | ||   |    | | | || || || |
Db     556 GGCTGCCGGGGCAGGAATGCAGNCCCTGGA 585


RESULT   6
AI313487
LOCUS       AI313487      498 bp    mRNA            EST       03-FEB-1999
DEFINITION  qp80f06.x1 Soares_fetal_lung_NbHL19W Homo sapiens cDNA clone
            IMAGE:1929347 3' similar to TR:O35347 O35347 DGCR6 PROTEIN
            ;contains PTR5.b1 MER22 repetitive element ;, mRNA sequence.
ACCESSION   AI313487
NID         g4019092
VERSION     AI313487.1  GI:4019092
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 498)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On May 18, 1995 this sequence version replaced gi:811067.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1108   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 481.
FEATURES             Location/Qualifiers
     source          1. .498
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1929347"
                     /clone_lib="Soares_fetal_lung_NbHL19W"
                     /dev_stage="19 weeks"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: lung; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCAATTTTTTTTTTTTTTTTT-3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by Bento
                     Soares and M.Fatima Bonaldo. This library was constructed
                     from the same fetus as the fetal heart library, Soares
                     fetal heart NbHH19W."
BASE COUNT      108 a    156 c    166 g     68 t
ORIGIN


  Query Match             2.8%;  Score 34.2;  DB 44;  Length 498;
  Best Local Similarity   50.3%;  Pred. No. 4.6;
  Matches   84;  Conservative    0;  Mismatches   83;  Indels    0;  Gaps    0;

Qy     850 CAAATCTCCCCTGAACATCAGGCTATTCTGTCGAAACGTCTTGAGTCTGTCGAATCCGAT 909
           || | |   ||||    ||||||   || |  |  |   || ||| | || |||  |
Db     316 CACAACCTGCCTGTGGTTCAGGCGGCTCAGCAGCGAGAACTAGAGGCCGTGGAACACCGG 375

Qy     910 ATTCGTCTTGAGCAGAATACCATGGATATGACCCGAATCGATGCGCGCAAGATGCAGATG 969
           || |||   ||||||    | |||||   ||     |||   | ||   |   | || ||
Db     376 ATCCGTGAGGAGCAGCGGGCGATGGACCAGAAGATCATCCTGGAGCTGGACCGGAAGGTG 435

Qy     970 ACGGGCGATCTGATTATGAAGAACTCGGTCACGGTCGGTGGTATTGC 1016
           | | | | | ||  |    | |   ||    ||| | |||  || |
Db     436 GCTGACCAGCAGAGCACACTGGAGAAGGCGGGGGTGGCTGGCTTTAC 482


RESULT   7
AA116428
LOCUS       AA116428      450 bp    mRNA            EST       17-FEB-1997
DEFINITION  mq08d09.r1 Soares 2NbMT Mus musculus cDNA clone IMAGE:578129 5',
            mRNA sequence.
ACCESSION   AA116428
NID         g1671440
VERSION     AA116428.1  GI:1671440
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 450)
```

```
DE   DKFZp547H0162)
XX
KW   EST; expressed sequence tag.
XX
OS   Homo sapiens (human)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria;
OC   Primates; Catarrhini; Hominidae; Homo.
XX
RN   [1]
RP   1-692
RA   Ottenwaelder B., Obermaier B., Mewes W., Gassenhuber J., Wiemann S.;
RT   ;
RL   Submitted (12-MAR-1999) to the EMBL/GenBank/DDBJ databases.
RL   MIPS, Am Klopferspitz 18a D-82152 Martinsried, GERMANY
XX
CC   Clone from S. Wiemann, sequenced by MediGenomix within the cDNA
CC   sequencing consortium of the German Genome Project
CC   No s1 sequence available
CC   This clone is available at the RZPD in Berlin
CC   Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
CC   Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de
XX
FH   Key             Location/Qualifiers
FH
FT   source          1. .692
FT                   /db_xref="taxon:9606"
FT                   /organism="Homo sapiens"
FT                   /clone="DKFZp547H0162"
FT                   /clone_lib="547 (synonym: hfbr1). Vector pAMP1; host
FT                   X1-2blue; sites NotI + SalI"
FT                   /dev_stage="fetal"
FT                   /tissue_type="brain"
XX
SQ   Sequence 692 BP; 207 A; 168 C; 150 G; 167 T; 0 other;
```

```
  Query Match              2.8%;  Score 34.8;  DB 53;  Length 692;
  Best Local Similarity   55.9%;  Pred. No. 3.4;
  Matches   66;  Conservative    0;  Mismatches   52;  Indels    0;  Gaps    0;

Qy   1113 CCGTGAAAGTTCACGTAAATCGACCAGCCTGATTCAGGAAATGCTGAAAACAATGGAGAG 1172
          || ||  | |||   |   | || | | | ||||    |   |  ||  || ||| ||
Db    347 CCATGGTATTTCCTCTTTTCCCACTATCTTCAATCATTCTAAAATTCAATGAAGGGACAG 288

Qy   1173 CATTAACCAGTCGAAAGCATCCGCACTCGCTGCTATCGCAGGCAATATTCGCGCTTAA 1230
          ||| |||   || ||| ||||  | ||  ||  || ||||   ||  | || ||||
Db    287 CATGAACTTTTCAAAACCATCTATATTCTTTGAAATGGCAGTTGATTATGGCACTTAA 230
```

```
RESULT   4
AI572641
LOCUS       AI572641      522 bp    mRNA            EST       13-MAY-1999
DEFINITION  tq68a08.x1 NCI_CGAP_Lu19 Homo sapiens cDNA clone IMAGE:2213942 3'
            similar to TR:O35347 O35347 DGCR6 PROTEIN ;contains PTR5.b1 PTR5
            repetitive element ;, mRNA sequence.
ACCESSION   AI572641
NID         g4536015
VERSION     AI572641.1  GI:4536015
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 522)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On May 7, 1998 this sequence version replaced gi:3121445.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov

            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 1105   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 484
            POLYA=No.
FEATURES              Location/Qualifiers
     source           1. .522
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:2213942"
                      /clone_lib="NCI_CGAP_Lu19"
                      /tissue_type="squamous cell carcinoma, poorly
                      differentiated (4 pooled tumors, including primary and
                      metastatic)"
                      /dev_stage="adult"
                      /lab_host="DH10B (phage-resistant)"
                      /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                      modified polylinker; 1st strand cDNA was prepared from
                      pooled lung tumor tissue, and was then primed with a Not I
                      - oligo(dT) primer. Double-stranded cDNA was ligated to
                      Eco RI adaptors (Pharmacia), digested with Not I and
                      cloned into the Not I  and Eco RI sites of the modified
                      pT7T3 vector.  Library went througb one round of
                      normalization. Library constructed by Bento Soares and M.
                      Fatima Bonaldo. "
BASE COUNT      114 a    161 c    174 g     73 t
ORIGIN
```

```
  Query Match              2.8%;  Score 34.6;  DB 48;  Length 522;
  Best Local Similarity   50.9%;  Pred. No. 3.5;
  Matches   82;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy    850 CAAATCTCCCCTGAACATCAGGCTATTCTGTCGAAACGTCTTGAGTCTGTCGAATCCGAT 909
          || | |   ||||    ||||||   || |  |  |   || ||| | || |||  |
Db    316 CACAACCTGCCTGTGGTTCAGGCGGCTCAGCAGCGAGAACTAGAGGCCGTGGAACACCGG 375

Qy    910 ATTCGTCTTGAGCAGAATACCATGGATATGACCCGAATCGATGCGCGCAAGATGCAGATG 969
          || |||   ||||||    | |||||   ||     |||   | ||   |   | || ||
Db    376 ATCCGTGAGGAGCAGCGGGCGATGGACCAGAAGATCATCCTGGAGCTGGACCGGAAGGTG 435

Qy    970 ACGGGCGATCTGATTATGAAGAACTCGGTCACGGTCGGTGG 1010
          | | | | | ||  |    | |   || |  ||| | |||
Db    436 GCTGACCAGCAGAGCACACTGGAGAAGGGCGGGGTGGCTGG 476
```

```
RESULT   5
AI245708
LOCUS       AI245708      597 bp    mRNA            EST       28-JAN-1999
DEFINITION  qk31g03.x1 NCI_CGAP_Kid3 Homo sapiens cDNA clone IMAGE:1870612 3'
            similar to TR:Q14129 Q14129 DGCR6. ;contains PTR5.b1 TAR1
            repetitive element ;, mRNA sequence.
ACCESSION   AI245708
NID         g3841105
VERSION     AI245708.1  GI:3841105
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 597)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
```

```
KEYWORDS    EST.
SOURCE      Leishmania major.
  ORGANISM  Leishmania major
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Leishmania.
REFERENCE   1  (bases 1 to 191)
  AUTHORS   Schneider,H.
  TITLE     Schneider,H. Unpublished (1996)
  JOURNAL   Unpublished (1996)
COMMENT     On Jan 25, 1995 this sequence version replaced gi:637859.

            Contact: Schneider H
            Laboratorio de Biologia Molecular
            Universidade Federal do Para
            Departamento de Genetica, Campus Universitario do Guama, Caixa
            Postal 8607, 66.075-900, Belem, Para, Brazil
            Tel: 55912111568
            Fax: 55912111568
            Email: horacio@cuxiu.cbio,ufpa.br/horacio@marajo.secom.ufpa.br
            Clones for these ESTs are available from:
             Blackwell JM
             Cambridge Institute for Medical Research
             Wellcome Trust/MRC Building
             Addenbrooke's Hospital, Hills Road,
             Cambridge CB2 2XY, UK
             Tel: 01223 336 143
             Fax: 01223 331 206
             Email: jmb37@cus.cam.ac.uk
            Seq primer: T3.
FEATURES              Location/Qualifiers
     source           1. .191
                      /organism="Leishmania major"
                      /strain="LV39"
                      /db_xref="taxon:5664"
                      /map="18"
                      /clone="lmk254"
                      /clone_lib="LmLV39cDNA"
                      /lab_host="XL1-Blue MRF'"
                      /note="Vector: Lambda UNIZAP; Site_1: NotI; Site_2: XhoI;
                      Mid or late log promastigotes.Full length : splice leader
                      oligo used for second strand synthesis, directional
                      cloning NotI at 5' end, XhoI at 3' end."
BASE COUNT       61 a     46 c     65 g     19 t
ORIGIN
```

```
  Query Match             3.0%;  Score 37.2;  DB 26;  Length 191;
  Best Local Similarity   59.4%;  Pred. No. 0.36;
  Matches   63;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy     239 ATACCGACATGAATCAGGATTTGAATGCTCTGGCAAATAATGTCACGACTAAAGCGAATG 298
           | ||| | ||| | |||||    || |  | |   || || | | || |  | |||||
Db      25 AGACCAAGATGGACCAGGAGACAAAGGACCAGATGAAAAACGCCGCGGCGGAGGCGAAGG 84

Qy     299 AGGTTGTGCAAACCCAGTTACGCGAGCAGCAGGCAGAAGTCGGAAA 344
           |   ||||||   | || | |  |||| | |||  || || || ||
Db      85 ACAATGTGCACGACAAGATCCAGGAGCTGAAGGACGATGTGGGGAA 130
```

```
RESULT   2
AI392302/c
LOCUS       AI392302      476 bp    mRNA            EST       03-FEB-1999
DEFINITION  NCSC1F2T3 Subtracted Conidial Neurospora crassa cDNA clone SC1F2 5'
            similar to pyruvate kinase, mRNA sequence.
ACCESSION   AI392302
NID         g4220109
VERSION     AI392302.1  GI:4220109
KEYWORDS    EST.
SOURCE      Neurospora crassa.
  ORGANISM  Neurospora crassa
            Eukaryota; Fungi; Ascomycota; Euascomycetes; Pyrenomycetes;
            Sordariales; Sordariaceae; Neurospora.
REFERENCE   1  (bases 1 to 476)
  AUTHORS   Nelson,M.A., Kang,S., Braun,E.L., Crawford,M.E., Dolan,P.L.,
            Leonard,P.M., Mitchell,J., Armijo,A.M., Bean,L., Blueyes,E.,
            Cushing,T., Errett,A., Fleharty,M., Gorman,M., Judson,K.,
            Miller,R., Ortega,J., Pavlova,I., Perea,J., Todisco,S.,
            Trujillo,R., Valentine,J., Wells,A., Werner-Washburne,M., Yazzie,S.
            and Natvig,D.O.
  TITLE     Expressed sequences from conidial, mycelial, and sexual stages of
            Neurospora crassa
  JOURNAL   Fungal Genet. Biol. 21, 348-363 (1997)
  MEDLINE   97435549
COMMENT     On Jan 14, 1998 this sequence version replaced gi:1797050.

            Contact: Natvig,D.O./Nelson,M.A.
            Department of Biology
            University of New Mexico
            Castetter Hall, Albuquerque, NM 87131, USA
            Tel: 505 277 3411
            Fax: 505 277 0304
            Email: ngp@biology.unm.edu.
FEATURES              Location/Qualifiers
     source           1. .476
                      /organism="Neurospora crassa"
                      /strain="74-OR23-IV A (FGSC 2489)"
                      /db_xref="taxon:5141"
                      /map="2"
                      /clone="SC1F2"
                      /clone_lib="Subtracted Conidial"
                      /sex="Mating Type A"
                      /tissue_type="Conidia"
                      /dev_stage="Germinating conidia"
                      /lab_host="E. coli"
                      /note="Vector: pBlueScript SK (-); Site_1: EcoRI; Site_2:
                      XhoI; mRNA isolated from germinating conidia, grown in 1x
                      Vogel's, 2%  sucrose for 4.5 hours.  cDNA directionally
                      cloned into pBluescript  SK(-) using the Uni-ZAP XR vector
                      system (Stratagene, La Jolla, CA).  Previously identified
                      highly expressed clones were subtracted from  this
                      library."
BASE COUNT      105 a    163 c     92 g    113 t      3 others
ORIGIN
```

```
  Query Match             2.9%;  Score 35.2;  DB 45;  Length 476;
  Best Local Similarity   51.6%;  Pred. No. 2.2;
  Matches   79;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy     513 GAATGCGTTGTCCGGTAGTATTTCCCAGAGCGCGCTTCAGTTGGGGATCACTGGCGTGGG 572
           | | | ||| ||  |  |||  |   || |||||     |  | | |||  ||  ||  |
Db     190 GGAGGTGTTATCGAGGTGTAGATTGAAGCGCGCGGGGTTGACGNGAATCTTTGAGGTATG 131

Qy     573 CGCCAAACTGGAATATAAGGGGCTGCAGAATGAAAGAGGCGCGCTTAAACATAATGCCGC 632
             |  || ||| || ||     ||| |||  | |   |     | || ||| |      ||
Db     130 GGGAAATCTGCAAGATCGACAGCTTCAGCGTCAGTAAATAAAGTTTCAACCTGGGCAAGC 71

Qy     633 GAAGATCGATAAACTGACCACTGAAAGCCACAG 665
           | |   ||    | |||| | || ||| |||||
Db      70 GCAATGCGCAGGAATGACAATTGCAAGACACAG 38
```

```
RESULT   3
HSM000122/c
ID   HSM000122  standard; RNA; EST; 692 BP.
XX
AC   AL035830;
XX
SV   AL035830.1
XX
NI   e1394102
XX
DT   12-MAR-1999 (Rel. 59, Created)
DT   12-MAR-1999 (Rel. 59, Last updated, Version 1)
XX
DE   Homo sapiens mRNA; EST DKFZp547H0162_r1  (from clone
```

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 1999, 17:13:44 ; Search time 1495.71 Seconds
(without alignments)
1622.117 Million cell updates/sec

Title: US-09-068-804-2
Perfect score: 1230
Sequence: 1 ATGTTAATTAGTAATGTGGG.....CAGGCAATATTCGCGCTTAA 1230

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 986266752 residues

Database : EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		37.2	3.0	191	26	W88389	W88389 lmEST0337 L
c 2		35.2	2.9	476	45	AI392302	AI392302 NCSC1F2T3
c 3		34.8	2.8	692	53	HSM000122	Al035830 Homo sapi
4		34.6	2.8	522	48	AI572641	AI572641 tq68a08.x
5		34.4	2.8	597	44	AI245708	AI245708 qk31g03.x
6		34.2	2.8	498	44	AI313487	AI313487 qp80f06.x
7		33.6	2.7	450	28	AA116428	AA116428 mq08d09.r
8		33.6	2.7	452	42	AI073883	AI073883 oo12c04.x
9		33.4	2.7	478	36	AA636805	AA636805 vr16e12.r
c 10		33	2.7	439	30	AA234615	AA234615 zs39d02.r
11		33	2.7	520	34	AA463473	AA463473 zx98e09.s
12		33	2.7	535	41	AI016764	AI016764 ov27d06.x
13		33	2.7	530	44	AI269888	AI269888 qk04c01.x
c 14		33	2.7	503	45	AI392307	AI392307 NCSM2G11T
15		33	2.7	655	47	AI488791	AI488791 EST247130
c 16		33	2.7	699	50	AI690567	AI690567 tq12h10.x
17		32.8	2.7	521	35	AA572440	AA572440 vl82a03.r
18		32.8	2.7	875	40	AA991144	AA991144 BSBmMFSZ0
c 19		32.8	2.7	567	49	AI664835	AI664835 605002B12
20		32.6	2.7	404	27	U37233	U37233 HSU37233 Hu
c 21		32.6	2.7	384	47	AI511740	AI511740 LD43242.3
c 22		32.4	2.6	338	53	HSM000128	Al035836 Homo sapi
c 23		32.2	2.6	438	29	AA131656	AA131656 zl33g05.r
24		32.2	2.6	570	47	AI483570	AI483570 EST249420
25		32.2	2.6	571	47	AI487346	AI487346 EST245668
26		32.2	2.6	466	47	AI488761	AI488761 EST247100
c 27		32	2.6	465	21	T69405	T69405 yc38b10.s1
c 28		32	2.6	388	24	N37829	N37829 19056 Lambd
29		32	2.6	521	36	C80865	C80865 C80865 Mous
c 30		32	2.6	468	45	AI324542	AI324542 mr86e09.y
c 31		31.8	2.6	358	21	R05847	R05847 ye88f10.s1
c 32		31.8	2.6	476	22	R55103	R55103 yg87d02.r1
33		31.8	2.6	445	26	W29444	W29444 mc04h10.r1
34		31.8	2.6	422	26	W62069	W62069 md84e05.r1
c 35		31.8	2.6	370	26	W84658	W84658 zh52e05.s1
c 36		31.8	2.6	273	26	W84698	W84698 zh52a05.s1
37		31.8	2.6	540	28	AA118064	AA118064 mn07h03.r
38		31.8	2.6	428	30	AA222554	AA222554 my16a08.r
c 39		31.8	2.6	414	35	AA584671	AA584671 no09a09.s
c 40		31.8	2.6	651	37	AA675935	AA675935 SWOvL3CAN
c 41		31.8	2.6	538	38	AA777615	AA777615 zi95a01.s
c 42		31.8	2.6	610	42	AI124356	AI124356 SWOvAFCAP
c 43		31.8	2.6	656	42	AI124369	AI124369 SWOvAFCAP
c 44		31.8	2.6	331	42	AI151695	AI151695 SWOvAFCAP
c 45		31.8	2.6	552	43	AI205372	AI205372 SWOvAFCAP

ALIGNMENTS

RESULT 1
W88389
LOCUS W88389 191 bp mRNA EST 08-DEC-1998
DEFINITION lmEST0337 LmLV39cDNA Leishmania major cDNA clone lmk254 5' END, mRNA sequence.
ACCESSION W88389
NID g1402412
VERSION W88389.1 GI:1402412

```
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 192849)
  AUTHORS   Sulston,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAY-1999) Nematode Sequencing Project, Sanger Centre,
            Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics,
            Washington University, St. Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     On May 28, 1999 this sequence version replaced gi:4107411.
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
            and the release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc. Order of segments is not known; 800 n's separate
            segments.
            * NOTE: This is a 'working draft' sequence.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
FEATURES             Location/Qualifiers
     source          1. .192849
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /chromosome="unknown"
                     /clone="Y60A3"
BASE COUNT    60465 a  33065 c  33694 g  61624 t   4001 others
ORIGIN


  Query Match             3.6%;  Score 37.6;  DB 34;  Length 192849;
  Best Local Similarity   50.6%;  Pred. No. 12;
  Matches   91;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy     504 TGATATTCTTTCCAAAATGGGAGGCTGGTTATTACCAGGTAAGGACGGTAATACCGTTAA 563
           ||  | |  ||  ||    ||||   |||||  |||||  ||   ||  | ||
Db  178791 TGTGACTGGTTTGAATGCAGGAGTACGGTTAAGACCAGAGAATTCCGAAATTAGTTCGGC 178732

Qy     564 GCTAGATGTTACCTCACTCAAAAATGATTTAAACAGTTTAGTCAATAAATATAATCAAAT 623
           | | || |||    .  ||| || | ||||||| | |||    || || || |||    |
Db  178731 GATTGAAGTTGGGATTTTCATAATTTATTTAAAAATTTTGAAAAAAAATTAGAATTTTTT 178672

Qy     624 AAACAGTAATACCGTTTTATTTCCAGCGCAGTCAGGCAGCGGCGTTAAAGTAGCCACTGA 683
              |||| | |   |||| | ||      |||      |   |    ||| | || | |||
Db  178671 TTTCAGTCAAAAATTTTTTTCTCTCAAAAAGTTCAAAAATCGTCAAAAACTTGCGATTGA 178612


Search completed: October  5, 1999, 18:01:01
Job time: 11504 sec
```

This Page Blank (uspto)

```
  AUTHORS   Brown,H.J. and Coppel,R.
  TITLE     Primary structure of a Plasmodium falciparum rhoptry antigen
  JOURNAL   Mol. Biochem. Parasitol. (1991) In press
FEATURES             Location/Qualifiers
     source          1. .4395
                     /organism="Plasmodium falciparum"
                     /db_xref="taxon:5833"
                     /cell_line="asexual culture"
     misc_feature    join(125. .188,339. .448,891. .1851,2051. .2113,2537. .3301,
                     3445. .4108)
                     /note="putative CDS; putative"
     exon            125. .188
     exon            339. .448
     exon            891. .1851
     exon            2051. .2113
     exon            2274. .2330
     exon            2537. .3301
     exon            3445. .4111
BASE COUNT     1731 a    452 c    523 g   1689 t
ORIGIN


  Query Match             3.7%;  Score 38.4;  DB 36;  Length 4395;
  Best Local Similarity   58.9%;  Pred. No. 4.5;
  Matches   66;  Conservative    0;  Mismatches   46;  Indels    0;  Gaps    0;

Qy     533 TATTACCAGGTAAGGACGGTAATACCGTTAAGCTAGATGTTACCTCACTCAAAAATGATT 592
           | ||   || |||   |||| |||  ||    | |    ||| || |   |||||  |||
Db    2316 TCTTTAAAGCTAAAATCGGTGATATTGTGGCTCAATTCTTTAACTGAAGGAAAAAAAATT 2257

Qy     593 TAAACAGTTTAGTCAATAAATATAATCAAATAAACAGTAATACCGTTTTATT 644
           ||| |  | | | ||||||||||   | ||| |   ||  | | |||||||
Db    2256 AAAATAAATAAATAAATAAATATATATATATATATCATAGCATCCTTTTATT 2205


RESULT  13
CEY102G3_2/c
WPCOMMENT
  Sequence split into 4 fragments  LOCUS CEY102G3 Accession AL020985
     Fragment Name      Begin     End
     CEY102G3_0             1     110000
     CEY102G3_1        100001     210000
     CEY102G3_2        200001     310000
     CEY102G3_3        300001     366054
Continuation (3 of 4) of CEY102G3 from base 200001 (AL020985 Caenorhabditis elegans chro


  Query Match             3.6%;  Score 37.6;  DB 34;  Length 110000;
  Best Local Similarity   50.6%;  Pred. No. 11;
  Matches   91;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy     504 TGATATTCTTTCCAAAATGGGAGGCTGGTTATTACCAGGTAAGGACGGTAATACCGTTAA 563
           ||  | |  ||  ||    ||||   |||||  |||||  ||   ||  | ||
Db   27962 TGTGACTGGTTTGAATGCAGGAGTACGGTTAAGACCAGAGAATTCCGAAATTAGTTCGGC 27903

Qy     564 GCTAGATGTTACCTCACTCAAAAATGATTTAAACAGTTTAGTCAATAAATATAATCAAAT 623
           | | || |||       ||| || | |||||||| | |||    || || || |||    |
Db   27902 GATTGAAGTTGGGATTTTCATAATTTATTTAAAAATTTTGAAAAAAAATTAGAATTTTTT 27843

Qy     624 AAACAGTAATACCGTTTTATTTCCAGCGCAGTCAGGCAGCGGCGTTAAAGTAGCCACTGA 683
           ||||  | |   ||||  | ||      |||    |   |    ||| | || | |||
Db   27842 TTTCAGTCAAAAATTTTTTTCTCTCAAAAAGTTCAAAAATCGTCAAAAACTTGCGATTGA 27783


RESULT  14
CEY113G7_0/c
WPCOMMENT
  Sequence split into 8 fragments  LOCUS CEY113G7 Accession AL031113
     Fragment Name      Begin     End
     CEY113G7_0             1     110000
     CEY113G7_1        100001     210000
     CEY113G7_2        200001     310000
     CEY113G7_3        300001     410000
     CEY113G7_4        400001     510000
     CEY113G7_5        500001     610000
     CEY113G7_6        600001     710000
     CEY113G7_7        700001     711232
LOCUS       CEY113G7   711232 bp    DNA             HTG       12-JAN-1999
DEFINITION  Caenorhabditis elegans chromosome V clone Y113G7, WORKING DRAFT
            SEQUENCE, in unordered pieces.
ACCESSION   AL031113
NID         g4033889
VERSION     AL031113.1  GI:4033889
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 711232)
  AUTHORS   Lennard,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JAN-1999) Nematode Sequencing Project, Sanger Centre,
            Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics,
            Washington University, St. Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     On Dec 19, 1998 this sequence version replaced gi:4029002.
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
            and the release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc. Order of segments is not known; 800 n's separate
            segments.
            * NOTE: This is a 'working draft' sequence.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
FEATURES             Location/Qualifiers
     source          1. .711232
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /chromosome="V"
                     /clone="Y113G7"
BASE COUNT    231695 a 124651 c 124135 g 229951 t    800 others
ORIGIN


  Query Match             3.6%;  Score 37.6;  DB 34;  Length 110000;
  Best Local Similarity   50.6%;  Pred. No. 11;
  Matches   91;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy     504 TGATATTCTTTCCAAAATGGGAGGCTGGTTATTACCAGGTAAGGACGGTAATACCGTTAA 563
           ||  | |  ||  ||    ||||   |||||  |||||  ||   ||  | ||
Db   86112 TGTGACTGGTTTGAATGCAGGAGTACGGTTAAGACCAGAGAATTCCGAAATTAGTTCGGC 86053

Qy     564 GCTAGATGTTACCTCACTCAAAAATGATTTAAACAGTTTAGTCAATAAATATAATCAAAT 623
           | | || |||       ||| || | |||||||| | |||    || || || |||    |
Db   86052 GATTGAAGTTGGGATTTTCATAATTTATTTAAAAATTTTGAAAAAAAATTAGAATTTTTT 85993

Qy     624 AAACAGTAATACCGTTTTATTTCCAGCGCAGTCAGGCAGCGGCGTTAAAGTAGCCACTGA 683
           ||||  | |   ||||  | ||      |||    |   |    ||| | || | |||
Db   85992 TTTCAGTCAAAAATTTTTTTCTCTCAAAAAGTTCAAAAATCGTCAAAAACTTGCGATTGA 85933


RESULT  15
CEY60A3/c
LOCUS       CEY60A3    192849 bp    DNA             HTG       27-MAY-1999
DEFINITION  Caenorhabditis elegans chromosome unknown clone Y60A3, WORKING
            DRAFT SEQUENCE, in unordered pieces.
ACCESSION   AL021574
NID         g4914474
VERSION     AL021574.2  GI:4914474
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
```

```
                AWGHDTLKPVSKSYSDWFDTGLTIVDGLDTAIIMGLEDEATEATEWIQNTLTFEKDRM
                VNFFECTIRVLGGMMSAFHLTGKKMFLEKSVDLGDRLLSAFKSPSPIPYSDVNLLKRT
                ATNPQWGADSSLSEVTTVQLEYRALSRASGNSTYEDLTFNVFKHIHKIGCETHEGLCG
                MFINANTGNFKKDSTITFGARSDSFYEYLFKQWIQTGKTIDWLKEDYGKAMKAMEKYL
                YRNSKPNKMFFIGELLSGQTYSPKMDHLVCFIAGTLSQGSQHGFPRKHLDMAEKIGET
                CHNMYDNPTGLGPEIAHFNMIPGKEDLYVKPLDAHCLLRPEAIEAWFYLYRFTGDKKY
                QEWGWSAFQAIEKYARIPTGGYSSISNVKQINVRFRDSMESFLLGETFKYLYLLLGDD
                QTVLPLDKWVFTTEGHPLPIYDH"
gene            complement(5114. .5545)
                /gene="T03G11.5"
CDS             complement(5114. .5545)
                /gene="T03G11.5"
                /note="Similar to exopolygalacturonase."
                /codon_start=1
                /evidence=not_experimental
                /protein_id="AAA82447.1"
                /db_xref="PID:g1086861"
                /db_xref="GI:1086861"
                /translation="MYHRSHHYNKQRGEIDHRPITTTTMSTLSTFTSTAITISPALAD
                VFSRQIEMSKIETSSDPCECEVKHCPQKCFTALVSGTTTALATTRPTTTATSSTFNNI
                TLPENSKESQLLVNSYPHVFQSCSLIIFFVTKMFIYKLFYQ"
gene            10378. .13805
                /gene="T03G11.3"
CDS             join(10378. .10411,11025. .11095,11149. .11280,11336. .11483,
                12239. .12311,12365. .12637,12683. .12790,13541. .13618,
                13667. .13805)
                /gene="T03G11.3"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="AAA82448.1"
                /db_xref="PID:g1086862"
                /db_xref="GI:1086862"
                /translation="MASFFNRRQREDSADPNEPTFPCPICDRRFIKSSLEKHESACRK
                LASLHRKPFDSGKQRASGSDLTYADIKKVQHEKNKNGGVFPRPQTNWRERHGNFIDAV
                SSSKRVDYALKTGAPLPPPPKTAVPSDYVQCEYCSRNFNAAAAERHIPFCREQATRKQ
                GGNLKSSGGNRALTGNYRSTPSKNEVRALEFMAIFFSFFWFQGKKQESSSRNGSAERK
                PTTRGRDGSLLRARRDDSNDITSRRKSLDTRTSLTTGQASNRHTSLSAGMRPTLPPMT
                PSSNSPHYPIPLFNCLIFISLLVYFFNGSRASTTTTASRSGSRTSSRACPRDDSRDSR
                LSQAKNNSRNNSRSRIF"
gene            complement(14074. .15488)
                /gene="T03G11.6"
CDS             complement(join(14074. .14208,14259. .14309,14355. .14532,
                14581. .14690,14970. .15056,15103. .15372,15429. .15488))
                /gene="T03G11.6"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="AAA82449.1"
                /db_xref="PID:g1086863"
                /db_xref="GI:1086863"
                /translation="MSFQLERTYGLRAQHWYQPETSLSDVAQNLFAASLPDGETQQFL
                DNSNKIASNFFWQTVRNLGVMILSTIYSKTDINGMTGFGNMFLFSENQFAKFLAIEKE
                SWNSTDKKVLDLGAGNGDITEHMRPFFEHVYATELSSKMRNRLSWKGYNVLSALEWAE
                TDVKFDLVTAFNLLDRHYSPGKLLNDLWNVARRSNCNVIVSLVLPVSHYVEFNPNGRS
                TRPDNYLNVAGRTYADHVLRWTRLPYLCEGDMNNSAYYLPDAIFLLKPIEPESPSSTI
                TQETTSVSQESTSHASHTDL"
gene            15617. .21887
                /gene="T03G11.2"
CDS             join(16617. .16778,16824. .17033,17209. .17477,17536. .17632,
                17681. .17850,18794. .18851,19651. .19902,21469. .21591,
                21759. .21887)
                /gene="T03G11.2"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="AAA82450.1"
                /db_xref="PID:g1086864"
                /db_xref="GI:1086864"
                /translation="MNTVSWCVLGLSACVFLGLICFIRMAIIYCSQPSMATKAIHPHV
                FKSMVIMQLSHYIFFITDNLAFRVPSTGFVTASCAAMEPNHYLKIIYLSQMWGTYWSM
                FSSLLFSIIRMIVYVTPTRHAEWSSYVFYTFIPLAFFSPFGITYFMYPAKGYCRQWGS
                PLGFGSVFISYTETYDEIRHDIFFIPLTLAVPLLIIINNFALIVMAVGYSKTLKRGQS
                LFKQSSQLSLILTSIAMTVPYITSGILTAVVISYPDVNAYVLFIRSPITDIAQSSIPI
                IFYLTHPLFRKKIQKVYVANVIPSTYRYIWLLTIKYQAANSTATKKVLYFKILVSDLK
                NEADSHRHTGKTRFFLSTFTSKVQLLDTRIVKSNHSRRFVSLPAHHRVSLFYTFHLFF
                ITTGKRLVAFDFLVRACFYVSISALVVCLFSTICMFFGVYTVNLWLDIFLIVANIVVA
                ILMMFGLYYDKASLLVPFIVSEVSFLCDFLTFLKTNCCF"
gene            complement(24505. .25236)
                /gene="T03G11.8"
CDS             complement(join(24505. .24635,24680. .24820,24872. .25020,
                25073. .25236))
                /gene="T03G11.8"
                /note="Similar to human leukocyte surface protein (weak)."
                /codon_start=1
                /evidence=not_experimental
                /protein_id="AAA82451.1"
                /db_xref="PID:g1086865"
                /db_xref="GI:1086865"
                /translation="MNISEEEITISISPNANPVQKPIGHQISLVCSIKKTDSNGEKPG
                MIWKKHGGLDRTGNVEVKKLDDYTLGLIIRNSSVEDSGVYYCQAQVGSKVYMNKMDVI
                VFEDIVFRDKQLHFGQVLATASVNISCEVSAKKDSVITYWTRHGKQILEGGKHKFYSR
                GSILEIQNYQPEQDAGQYTCEVNEKYRIKTCKLS"
gene            34688. .38310
                /gene="T03G11.1"
CDS             join(34688. .35008,35080. .35400,35445. .35550,35596. .35764,
                35942. .36116,36785. .36906,36949. .37087,37130. .37340,
                37927. .38089,38142. .38310)
                /gene="T03G11.1"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="AAA82452.1"
                /db_xref="PID:g1086866"
                /db_xref="GI:1086866"
                /translation="MESTNQSHQNIDSMYHIDQPQQVVFQANLQNRSNRVMQQTPNNS
                SNHQNHLSLKQPHLSMNSTQLQSGFQTGNRMMPSFQVNPTTQDQVPHLSHMDQPPPTQ
                HSMNSISQSQQHPSQLSQLSCNSNQMSVQNHVLSPSSTNSNMSLNMRSNSISSSSSSP
                IKNSILPSTSPCTIPMNANSTTQLLMALNNNSPITSEQLQEIQNMTTISADGTEILLM
                QEPRLKQYPSIAEMCTTDLRSACKKRSLPHTGPKPRLCERLRGHENEVLEERNRQNMK
                EYTDRHQMYEAQAATLKAYQAKKALLQVQSSPDMGTSNVYGARSNSSQCISSNENSMS
                STSVEGPPAKKKRISKPKKALQSRLEKQTTITVAPQNSLTIPADSCNQGSTTTNAIRT
                LVNNTNLMGGSNTQNGNTMVNGQSSQNSSSNFFGDMNQTGTGFFMKDNSCFSHLGSGC
                RLISASSNDDSQNTFGGTGRAFQAAVPPPPENSQLQFQNFGMTETIPQQSNEHVSQQQ
                QQIAQSRPMNLNQQMFKTLRYGIRKSWGILRLTKQGLKIRYSTFKMVFELFCCQFIKS
                HFFYQVCVLLLQASPCVADQAGPSPSMPVSDPIATTSGSPVVVSAVAESTSESAELMD
                LETMQNL"
BASE COUNT    12418 a    6742 c    6538 g    12757 t
ORIGIN


  Query Match             3.8%;  Score 39.2;  DB 36;  Length 38455;
  Best Local Similarity   54.1%;  Pred. No. 3.6;
  Matches   80;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy     786 AAAAATGGTTCAGGATATTGATGGTTTAGGCGCGCCGGGAAAAGACTCAAAACTCGAAAT 845
           | |||||| || | | | ||| | | || | ||| | ||     ||||| ||||
Db   11068 ACAAATGGGACAAGGAAAAGTTGGCTCATTCGGGTCGGCAGAATCTAGAAAACAGGAAAA 11009

Qy     846 GGATAACGCCAAATATCAAGCCTGGCAGTCGGGTTTTAAAGCGCAGGAAGAAAATATGAA 905
           | ||   |  ||||||     ||  | |  | |  | |||   |  |  | || |     .
Db   11008 GAATTGTGATGAATATCCGAAATGAGATTGTGTTCATCAAGAAGATTACAATAAAAACCT 10949

Qy     906 AACCACATTACAGACGCTGACGCAAAAA 933
           | ||| ||| ||||||   |||  || |
Db   10948 ACCCATATTTCAGACGACAACGACAATA 10921


RESULT 12
PFARHOPH3/c
LOCUS       PFARHOPH3     4395 bp    DNA             INV       15-NOV-1992
DEFINITION  Plasmodium falciparum RhopH3 gene, putative CDS.
ACCESSION   M65059
NID         g160653
VERSION     M65059.1  GI:160653
KEYWORDS    .
SOURCE      Plasmodium falciparum DNA.
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 4395)
```

```
            *     2672     2680: gap of unknown length
            *     2681     4882: contig of 2202 bp in length
            *     4883     4891: gap of unknown length
            *     4892    10010: contig of 5119 bp in length
            *    10011    10019: gap of unknown length
            *    10020    16339: contig of 6320 bp in length
            *    16340    16348: gap of unknown length
            *    16349    18892: contig of 2544 bp in length
            *    18893    18901: gap of unknown length
            *    18902    21520: contig of 2619 bp in length
            *    21521    21529: gap of unknown length
            *    21530    40161: contig of 18632 bp in length
            *    40162    40170: gap of unknown length
            *    40171    43627: contig of 3457 bp in length
            *    43628    43636: gap of unknown length
            *    43637    48895: contig of 5259 bp in length
            *    48896    48904: gap of unknown length
            *    48905    54558: contig of 5654 bp in length
            *    54559    54567: gap of unknown length
            *    54568    58682: contig of 4115 bp in length
            *    58683    58691: gap of unknown length
            *    58692    63982: contig of 5291 bp in length
            *    63983    63991: gap of unknown length
            *    63992    74805: contig of 10814 bp in length
            *    74806    74814: gap of unknown length
            *    74815    85448: contig of 10634 bp in length
            *    85449    85457: gap of unknown length
            *    85458    94227: contig of 8770 bp in length
            *    94228    94236: gap of unknown length
            *    94237    99321: contig of 5085 bp in length
            *    99322    99330: gap of unknown length
            *    99331   104580: contig of 5250 bp in length
            *   104581   104588: gap of unknown length
            *   104589   106581: contig of 1993 bp in length
            *   106582   106589: gap of unknown length
            *   106590   110735: contig of 4146 bp in length
            *   110736   110743: gap of unknown length
            *   110744   114618: contig of 3875 bp in length
            *   114619   114626: gap of unknown length
            *   114627   121943: contig of 7317 bp in length
            *   121944   121951: gap of unknown length
            *   121952   125293: contig of 3342 bp in length
            *   125294   125301: gap of unknown length
            *   125302   129100: contig of 3799 bp in length
            *   129101   129108: gap of unknown length
            *   129109   142458: contig of 13350 bp in length
            *   142459   142466: gap of unknown length
            *   142467   148151: contig of 5685 bp in length
            *   148152   148159: gap of unknown length
            *   148160   152637: contig of 4478 bp in length.
FEATURES             Location/Qualifiers
     source          1. .152637
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /clone="Y34B4"
BASE COUNT    48440 a  26834 c  26549 g  48692 t   2122 others
ORIGIN


  Query Match             3.8%;  Score 39.2;  DB 34;  Length 152637;
  Best Local Similarity   54.1%;  Pred. No. 4.4;
  Matches   80;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy    786 AAAAATGGTTCAGGATATTGATGGTTTAGGCGCGCCGGGAAAAGACTCAAAACTCGAAAT 845
          | |||||| || |  |  | ||| | |  || | ||| | ||     |||||  ||||
Db  51220 ACAAATGGGACAAGGAAAAGTTGGCTCATTCGGGTCGGCAGAATCTAGAAAACAGGAAAA 51161

Qy    846 GGATAACGCCAAATATCAAGCCTGGCAGTCGGGTTTTAAAGCGCAGGAAGAAAATATGAA 905
          | ||   |   ||||||     ||  | |  | |  | |||   |  |  | || |
Db  51160 GAATTGTGATGAATATCCGAAATGAGATTGTGTTCATCAAGAAGATTACAATAAAAACCT 51101

Qy    906 AACCACATTACAGACGCTGACGCAAAAA 933
          | ||| ||| ||||||   |||  || |
```

```
Db  51100 ACCCATATTTCAGACGACAACGACAATA 51073


RESULT  11
CELT03G11/c
LOCUS       CELT03G11   38455 bp    DNA             INV       30-NOV-1995
DEFINITION  Caenorhabditis elegans cosmid T03G11.
ACCESSION   U41272
NID         g1086859
VERSION     U41272.1  GI:1086859
KEYWORDS    .
SOURCE      Caenorhabditis elegans strain=Bristol N2.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 38455)
  AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
  TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
  JOURNAL   Nature 368 (6466), 32-38 (1994)
  MEDLINE   94150718
REFERENCE   2  (bases 1 to 38455)
  AUTHORS   Geisel,C and Gattung,S.
  TITLE     The sequence of C. elegans cosmid T03G11
  JOURNAL   Unpublished (1995)
REFERENCE   3  (bases 1 to 38455)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-NOV-1995) Robert Waterston
COMMENT                 Submitted by:
                        Genome Sequencing Center
                        Department of Genetics, Washington University,
                        St. Louis, MO  63110, USA, and
                        Sanger Centre, Hinxton Hall
                        Cambridge CB10 IRQ, England
                        e-mail: rw@nematode.wustl.edu and jes@sanger.ac.uk
                        NEIGHBORING COSMID INFORMATION:
            The 5' cosmid is EGAP5, 200 bp overlap;3' end lies in a gap
            followed by the cosmid is F31A6.  Actual start of this cosmid is at
            base position 1 of CELT03G11; actual end is at 38455 of CELT03G11

            NOTES:
            Coding sequences below are predicted from computer analysis, using
            the program Genefinder(P. Green and L. Hillier, ms in preparation).
FEATURES             Location/Qualifiers
     source          1. .38455
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /map="X"
     gene            1786. .4889
                     /gene="T03G11.4"
     CDS             join(1786. .1881,1928. .2066,2331. .2437,2957. .3035,
                     3230. .3458,3506. .3717,3761. .3975,4035. .4223,4270. .4339,
                     4630. .4889)
                     /gene="T03G11.4"
                     /note="Similar to man(9)-alpha-mannosidase."
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAA82446.1"
                     /db_xref="PID:g1086860"
                     /db_xref="GI:1086860"
                     /translation="MRLVRDDPGLPLYNITSPKGHTFRRIRARARHWFSTRNTPTKAC
                     IIFTILLSIFVLITICSQGEAKREQPIVEVGYSKRPTNDRQKAVVKAFQHAWIGYKKY
```

```
                /protein_id="AAD19783.1"
                /db_xref="PID:g4388828"
                /db_xref="GI:4388828"
                /translation="MYSNSVGEFREGEIDGKLFLIYTLSSQTESLSSGEAAVDSDGGD
                GLSLEPRILCPTFRLKVKKLKPYTRDSDRSFFPLVVSPHFPTIQSEPATSYYVCVTCG
                DQFHKECIGSPLEFKHPSYPSLSLQLYSPPSDRVFCSCCQKRIDGLNYYSPTNNFTLH
                LFCAFKPIPFVIDHPKRHPHPLTIFPKQSFLPCHVCSLIKGHIPTYVCVRCVFVVHQD
                CIYFPYVIKISRHHHRISYTSSLPSGKWSCGVCRQMVDNDYGAYSCNKCDDYFVHSRC
                ALRRDLWDGIELEGVPEEPEIVVEPFITISDGIILHFSHGHHLKLKISRVYDEKMVCQ
                ACILPIYEGGYYSCMDECDFILHETCANAPCKKHHALHPHPLTLTVVTSEYEDNIGRF
                CCKACQRESCGFVYEDLRGKIGNYRRNKFQLDLRCASVSEPFEYLGHKHLLYLALNPE
                EEKLAICQICHEREDESSHCRKLNCIECDFVICFRCATLPYKARYQHDKHFLKFYETK
                EANDHSEWWFYSCDDCCTTLHIDCLLGEDMYMKPGHTIMYNMTGFDEHYEKKLRILRN
                NTLSRPCCSTCGQRCRQKIVFKYKEHHIFCAISCQKHVINYS"
repeat_region   complement(8141. .8239)
                /rpt_family="(TAAA)n"
repeat_region   complement(8562. .8620)
                /rpt_family="AT_rich"
repeat_region   complement(10082. .10132)
                /rpt_family="AT_rich"
repeat_region   10605. .10794
                /rpt_family="AT_rich"
mRNA            join(<10926. .12014,12058. .>12570)
                /gene="F9B22.3"
gene            <10926. .>12570
                /gene="F9B22.3"
                /note="F17L24.2; contains weak PHD-finger motifs"
CDS             join(10926. .12014,12058. .12570)
                /gene="F9B22.3"
                /codon_start=1
                /product="putative zinc-finger protein"
                /protein_id="AAD19772.1"
                /db_xref="PID:g4388817"
                /db_xref="GI:4388817"
                /translation="MNSNSVGAFREREIDGKPFYIVTFSSQTDTTIPSSSGETAIDGD
                HDLSFQQPLFLCPSLRLKADEENYSYSELLPFNISPLFPSISSADQQVQHLLDSDSRD
                ANICKLPVVPIYWCNNKHGDYAQFICRACDTFKIDTDYYICLTCDQMFHKECVESPLE
                LKNPSYQKPCVSCMCCDLALFSLFYHCPTYELSMHPVCAMKPIPIFIDNSKRHPHLLT
                FFPKQASLLCDVCGLIKEAFPTYVCPRCIFVAHQDCIYFPYVIKIYRHQHRISFTSSL
                PSGKWSCGVCRRKVDNNCGAYTCNKCSHYFVHTRCALRNDVWDGRDLEGVPEEPEIIL
                ELFETIADGIILHFSHDHHQGHFSDGTIVHPLFLALDPDQELEATCQICQQKATFLQK
                LNCIECDYVICFKCATLPYEAKYKHDTHLLTFWEGKEESDQYGWCEVCENQIVYSRED
                GFYACDDCCTTLHVHCLLGEDMYMLKPSQTITYIKREVHILPNNTRSRPLCHYHVDNR
                CPQKVVFKSEDMTFCSFRCFTRNAV"
repeat_unit     13153. .13559
                /note="LTRA"
misc_feature    13153. .18324
                /note="putative retroelement. Structure: long terminal
                repeat A (LTRA), putative pol polyprotein, long terminal
                repeat B (LTRB)"
mRNA            complement(<13798. .>17805)
                /gene="F9B22.4"
gene            complement(<13798. .>17805)
                /gene="F9B22.4"
                /note="F17L24.1; related to Tal-2 protein 2 (GB: X53976)"
CDS             complement(13798. .17805)
                /gene="F9B22.4"
                /codon_start=1
                /product="putative pol polyprotein with a Zn-finger CCHC
                type domain (prosite:QDOC50158) and a DDE integrase
                signature motif"
                /protein_id="AAD19773.1"
                /db_xref="PID:g4388818"
                /db_xref="GI:4388818"
                /translation="MAHLSVIGLKDVVLGTTPKPLTEEEEEDPEKRKKRDADEVARLE
                RCDKAKNVIFLNVADKVLRKIELCKTAAEAWETLDRLFMIRSLPHRVYTQLSFYTFKM
                QENKKIDENIDDFLKIVADLNHLQIDVTDEVQAILLLSSLPARYDGLVETMKYSNSRE
                KLRLDDVMVAARDKERELSQNNRPVVEGHFARGRPDGKNNNQGNKGKNRSRSKSADGK
                RVCWICGKEGHFKKQCYKWIERNKSKQQGSDNGESSLAKSTEAFNPAMVLLATDETLV
                VTDSIANEWVLDTGCSFHMTPRKDWFKDFKELSSGYVKMGNDTYSPVKGIGSIKIRNS
                DGSQVILTDVRYMPNMTRNLISLGTLEDRGCWFKSQDGILKIVKGCSTILKGQKRDTL
                YILDGVTEEGESHSSAEVKDETALWHSRLGHMSQKGMEILVKKGCLRREVIKELEFCE
                DCVYGKQHRVSFAPAQHVTKEKLAYVHSDLWGSPHNPASLGNSQYFISFVDDYSRKVW
                IYFLRKKDEAFEKFVEWKKMVENQSDRKVKKLRTDNGLEYCNHYFEKFCKEEGIVRHK
                TCAYTPQQNGIAERLNRTIMDKVRSMLSRSGMEKKFWAEAASTAVYLINRSPSTAINF
                DLPEEKWTGALPDLSSLRKFGCLAYIHADQGKLNPRSKKGIFTSYPEGVKGYKVWVLE
                DKKCVISRNVIFREQVMFKDLKGDSQNTISESDLEDLRVNPDMNDQEFTDQGGATQDN
                SNPSEATTSHNPVLNSPTHSQDEESEEEDSDAVEDLSTYQLVRDRVRRTIKANPKYNE
                SNMVGFAYYSEDDGKPEPKSYQEALLDPDWEKWNAAMKEEMVSMSKNHTWDLVIKPEK
                VKLIGCRWVFTRKAGIPGVEAPRFIARLVAKGFTQKEGVDYNEIFSPVVKHVSIRYLL
                SMVVHYNMELQQMDVKTAFLHGFLEEEIYMAQPEGFEIKRGSNKVCLLKRSLYGLKQS
                PRQWNLRFDEFMRGIKYTRSAYDSCVYFKKCNGDTYIYLLLYVDDMLIASANKSEVNE
                LKQLLSREFEMKDLGDAKKILGMEISRDRDAGLLTLSQEGYVKKVLRSFQMDNAKPVS
                TPLGIHFKLKAATDKEYEEQFERMKIVPYANTIGSIMYSMIGTRPDLAYSLGVISRFM
                SKPLKDHWQAVKWVLRYMRGTEKKKLCFRKQEDFLLRGYCDSDYGSNFDTRRSITGYV
                FTVGGNTISWKSKLQKVVAISSTEAEYMALTEAVKEALWLKGFAAELGHSQDYVEVHS
                DSQSAITLAKNSVHHERTKHIDIRLHFIRDIICAGLIKVVKIATECNPANIFTKTVPL
```

```
Query Match             3.8%;  Score 39.2;  DB 8;  Length 95765;
Best Local Similarity   50.4%;  Pred. No. 4.1;
Matches  123;  Conservative    0;  Mismatches  118;  Indels    3;  Gaps    1;

Qy    386 TTTCTGATGCTGAGATTTGGGATATGGTTTCCCAAAATATATCGGCGATAGGTGACAGCT 445
          |||| ||| | || ||        ||  ||     |   || ||   | || |||   |  |
Db  49860 TTTCAGATTCGGATATCCATTCGATTTTTGGATATTTTAGATATTCAATTGGTATAATAT 49919

Qy    446 ATCTGGGCGTTTATGAAAACGTTGTCGCAGTCTATACCGATTTTTATCAGGCCTTCAGTG 505
          |  | ||      |     || | | |    || |||   ||  ||| |       || |||
Db  49920 AGTTTGGTACAAAATTTAAGGATATACTCGTGTATTTTGAAATTTTTGTTCGATTTAGTT 49979

Qy    506 ATATTCTTTCCA---AAATGGGAGGCTGGTTATTACCAGGTAAGGACGGTAATACCGTTA 562
          | ||| |||| |   |  |  ||   |||||| || |  ||   |    |||| |   |
Db  49980 AGATTTTTTCTAGTTATGTTTGATTTTGGTTATAACTATCTAGTCATAAAAATAACAAAA 50039

Qy    563 AGCTAGATGTTACCTCACTCAAAAATGATTTAAACAGTTTAGTCAATAAATATAATCAAA 622
          |  || |||| |    | | |||  | |||| || | |  ||    |||| | ||  ||
Db  50040 AAATATATGTAAAAAAATTGAAAGCTAATTTCAAAATTGCAGATTCTAAAAAGAAAAAAG 50099

Qy    623 TAAA 626
            |||
Db  50100 GAAA 50103
```

```
RESULT  10
AC006734/c
LOCUS       AC006734   152637 bp    DNA              HTG       25-FEB-1999
DEFINITION  Caenorhabditis elegans clone Y34B4, WORKING DRAFT SEQUENCE, 26
            unordered pieces.
ACCESSION   AC006734
NID         g4309802
VERSION     AC006734.3  GI:4309802
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 152637)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Caehorhabditis elegans clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 152637)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-FEB-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Mar 1, 1999 this sequence version replaced gi:4263430.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 26 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     2671: contig of 2671 bp in length
```

```
  Matches   23;  Conservative  216;  Mismatches  173;  Indels    0;  Gaps    0;

Qy     606 CAATAAATATAATCAAATAAACAGTAATACCGTTTTATTTCCAGCGCAGTCAGGCAGCGG 665
           | ||  ||  ||  |  |||| ||  | |    |||  | | :: : ::  ::: :: ::
Db    1475 CTATCTATGCAAGTAGTTAAAGAGATAGAAGAATTTGGTACRRRRRRRRRRRRRRRRRRR 1416

Qy     666 CGTTAAAGTAGCCACTGAAGCGGAAGCGAGACAGTGGCTCAGTGAATTGAATTTACCGAA 725
           :  :::: ::  :  :::: ::::: :::: :: ::   :: :::  :::   :  :::
Db    1415 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1356

Qy     726 TAGCTGCCTGAAATCTTATGGATCCGGTTATGTCGTCACCGTTGATCTGACGCCATTACA 785
           ::  :   ::::    : :::   ::  : :  :  :  :  ::   :: :  :  : :
Db    1355 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1296

Qy     786 AAAAATGGTTCAGGATATTGATGGTTTAGGCGCGCCGGGAAAAGACTCAAAACTCGAAAT 845
           ::::: ::   :::: :  :: ::   ::: : :  :::::::::   ::::   ::::
Db    1295 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1236

Qy     846 GGATAACGCCAAATATCAAGCCTGGCAGTCGGGTTTTAAAGCGCAGGAAGAAAATATGAA 905
           ::: :: :  ::: :  :::   :: ::   :::    :::: : ::::::::: : :::
Db    1235 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1176

Qy     906 AACCACATTACAGACGCTGACGCAAAAATATAGCAATGCCAATTCATTGTACGACAACCT 965
           ::  : :  : ::: :  :: : ::::: : :: :: :  ::   :  : : :: ::
Db    1175 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1116

Qy     966 GGTAAAAGTGCTGAGCAGTACGATAAGTAGCAGCCTGGAAACCGCCAAAAGC 1017
           :: ::::: :  ::: :: : :: ::: :: ::   :::::  :  :::| |
Db    1115 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRATC 1064


RESULT   9
ATAC006528
LOCUS       ATAC006528  95765 bp    DNA             PLN       10-MAR-1999
DEFINITION  Arabidopsis thaliana chromosome II BAC F9B22 genomic sequence,
            complete sequence.
ACCESSION   AC006528
NID         g4388816
VERSION     AC006528.3  GI:4388816
KEYWORDS    HTG.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
            Arabidopsis.
REFERENCE   1  (bases 1 to 95765)
  AUTHORS   Lin,X., Kaul,S., Shea,T.P., Fujii,C.Y., Shen,M., VanAken,S.E.,
            Barnstead,M.E., Mason,T.M., Bowman,C.L., Ronning,C.M., Benito,M.,
            Carrera,A.J., Creasy,T.H., Buell,C.R., Town,C.D., Nierman,W.C.,
            Fraser,C.M. and Venter,J.C.
  TITLE     Arabidopsis thaliana chromosome II BAC F9B22 genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 95765)
  AUTHORS   Lin,X. and Kaul,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-FEB-1999) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org
REFERENCE   3  (bases 1 to 95765)
  AUTHORS   Lin,X.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAR-1999) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
COMMENT     On Mar 10, 1999 this sequence version replaced gi:4263537.
            Address all correspondence to:
            Xiaoying Lin
            The Institute for Genomic Research
            9712 Medical Center Dr.
            Rockville, MD 20850, USA
            e-mail: xlin@tigr.org
            BAC clone F9B22 is from Arabidopsis chromosome II and is near the
            molecular marker mi398.
            The orientation of the sequence is from SP6 to T7 end of the BAC
            clone.
            Genes were identified by a combination of three methods: Gene
            prediction programs including GRAIL (available by anonymous ftp
            from arthur.epm.ornl.gov), Genefinder (Phil Green, University of
            Washington), Genscan (Chris Burge,
            http://gnomic.stanford.edu/~chris/GENSCANW.html), and NetPlantGene
            (http://www.cbs.dtu.dk/netpgene/cbsnetpgene.html), searches of the
            complete sequence against a peptide database and the Arabidopsis
            EST database at TIGR (http://www.tigr.org/tdb/at/at.html).
            Annotated genes are named to indicate the level of evidence for
            their annotation. Genes with similarity to other proteins are named
            after the database hits. Genes without significant peptide
            similarity but with EST similarity are named as 'unknown' proteins.
            Genes without protein or EST similarity, that are predicted by more
            than two gene prediction programs over most of their length are
            annotated as 'hypothetical' proteins. Genes encoding tRNAs are
            predicted by tRNAscan-SE (Sean Eddy,
            http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are
            identified by repeatmasker (Arian Smit,
            http://ftp.genome.washington.edu/RM/RepeatMasker.html). Regions of
            genomic sequence that are not annotated as genes but have predicted
            exons by GRAIL are annotated as misc features.
FEATURES             Location/Qualifiers
     source          1. .95765
                     /organism="Arabidopsis thaliana"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="II"
                     /map="mi398"
                     /clone="F9B22"
     misc_feature    complement(1. .14531)
                     /note="overlap with BAC clone F17L24 (AC006218:1. .14531)."
     mRNA            join(<5. .271,371. .>1984)
                     /gene="F9B22.1"
     gene            <5. .>1984
                     /gene="F9B22.1"
                     /note="F17L24.4; contains weak PHD-finger motifs"
     CDS             join(<5. .271,371. .1984)
                     /gene="F9B22.1"
                     /codon_start=1
                     /product="putative zinc-finger protein, 5' partial"
                     /protein_id="AAD19788.1"
                     /db_xref="PID:g4388833"
                     /db_xref="GI:4388833"
                     /translation="SNPVGVFREREIDGKPFHIYTLSSQIDTTIPSSSGETAIDGDDH
                     LSLQQPLFLCPSLRLRVHEENNSYTEFLPYKTSPHFPSIGSVDQQLPVVPIYWCNSKE
                     LDYDQFNCGACERSKIGTNYYVCLTCDLMFHKECVESPLEIKHPSYPFLSLKLYNRPG
                     HRVSCMCCDLALCSMLYHCPTYKLSMHPVCAMKPIPIFIDHSKRHPHLLTFFPKQASL
                     LCDVCGLVKESFPTYVCVRCIFVVHRDCIYFPFVIKICRHHHRISFTSSLPSGKWSCG
                     VCRQKVDNNCGAYTCNKCNHYFVHTRCALREDVWDGKDLEGVPEEPEIILEPFETIAD
                     GIILHFSHGHHLTLEISRVYDENKFCQACILPIYEGNYYSCMDDCDFILHEACANASR
                     KKHHALHGHPLTLKIVPGEQEYYKDYFQCSACYRQSCGYVYEAGYYRIDLRCASVSEP
                     FQYQGHEHPLFLALDRQQELEATCQICQQKATFLRKLNCIECDYVICFKCATLPYEAK
                     YKHDKHFLTFREGKEESDQLDWCEVCEKRIVYSRKDGFYACDHCCTTLHVHCLLGEDM
                     YMLKPSQTITYIKREVHILPNNTRSRPFCHHHDEDRCPHKVVFKWQDQYTIFCSYRCL
                     TK"
     repeat_region   complement(336. .359)
                     /rpt_family="(CAT)n"
     repeat_region   3809. .3856
                     /rpt_family="AT_rich"
     repeat_region   4241. .4353
                     /rpt_family="(TAA)n"
     repeat_region   4267. .4392
                     /rpt_family="(TAAA)n"
     mRNA            join(<4613. .4880,5007. .6287,6366. .>6643)
                     /gene="F9B22.2"
     gene            <4613. .>6643
                     /gene="F9B22.2"
                     /note="F17L24.3; contains weak PHD-finger motifs"
     CDS             join(4613. .4880,5007. .6287,6366. .6643)
                     /gene="F9B22.2"
                     /codon_start=1
                     /product="putative zinc-finger protein"
```

```
                /db_xref="taxon:2661"
                /clone="HB101/pHS[4108,5357]."
     CDS        <1. .549
                /note="ORF"
                /codon_start=1
                /transl_table=11
                /protein_id="AAA98422.1"
                /db_xref="PID:g155328"
                /db_xref="GI:155328"
                /translation="KLVKKILSVFNLKQRFPQKNFGCPVNINKIRDSVIDKIKDSNSG
                NQLFCWMSQERTTYVSSMINRSIDEMAIHNGVVLTSDNKKNIFAAIEKKFPDIKLDEK
                SAQTSISHTALNEIASSGLRAKILKRYSSDMDLFNTQMKDLTNLVSSSVYDKIFNEST
                KVLQIEISAEVLKAVYRQSNTN"
     -10_signal 556. .561
                /gene="ippI"
     gene       556. .1073
                /gene="ippI"
     RBS        595. .600
                /gene="ippI"
     CDS        606. .1073
                /gene="ippI"
                /codon_start=1
                /transl_table=11
                /product="antigen"
                /protein_id="AAA98423.1"
                /db_xref="PID:g155329"
                /db_xref="GI:155329"
                /translation="MSLNITENESISTAVIDAINSGATLKDINAIPDDMMDDIYSYAY
                DFYNKGRIEEAEVFFRFLCIYDFYNVDYIMGLAAIYQIKEQFQQAADLYAVAFALGKN
                DYTPVFHTGQCQLRLKAPLKAKECFELVIQHSNDEKLKIKAQSYLDAIQDIKE"
     gene       1065. .2821
                /gene="ipaB"
     RBS        1065. .1070
                /gene="ipaB"
     CDS        1079. .2821
                /gene="ipaB"
                /codon_start=1
                /transl_table=11
                /product="antigen"
                /protein_id="AAA98424.1"
                /db_xref="PID:g155330"
                /db_xref="GI:155330"
                /translation="MHNVSTTTTGFPLAKILTSTELGDNTIQAANDAANKLFSLTIAD
                LTANQNINTTNAHSTSNILIPELKAPKSLNASSQLTLLIGNLIQILGEKSLTALTNKI
                TAWKSQQQARQQKNLEFSDKINTLLSETEGLTRDYEKQINKLKNADSKIKDLENKINQ
                IQTRLSELDPESPEKKKLSREEIQLTIKKDAAVKDRTLIEQKTLSIHSKLTDKSMQLE
                KEIDSFSAFSNTASAEQLSTQQKSLTGLASVTQLMATFIQLVGKNNEESLKNDLALFQ
                SLQESRKTEMERKSDEYAAEVRKAEELNRVMGCVGKILGALLTIVSVVAAAFSGGASL
                ALAAVGLALMVIDAIVQAATGNSFMEQALNPIMKAVIEPLIKLLSDAFTKMLEGLGVD
                SKKAKMIGSILGAIAGALVLVAAVVLVATVGKQAAAKLAENIGKIIGKTLTDLIPKFL
                KNFSSQLDDLITNAVARLNKFLGAAGDEVISKQIISTHLNQAVLLGESVNSATQAGGS
                VASAVFQNSASTNLADLTLSKYQVEQLSKYISEAIEKFGQLQEVIADLLASMSNSQAN
                RTDVAKAILQQTTA"
     gene       2723. .3932
                /gene="ipaC"
     -35_signal 2723. .2728
                /gene="ipaC"
     -10_signal 2745. .2750
                /gene="ipaC"
     RBS        2829. .2834
                /gene="ipaC"
     CDS        2841. .3932
                /gene="ipaC"
                /codon_start=1
                /transl_table=11
                /product="antigen"
                /protein_id="AAA98425.1"
                /db_xref="PID:g155331"
                /db_xref="GI:155331"
                /translation="MEIQNTKPTQTLYTDISTKQTQSSSETQKSQNYQQIAAHIPLNV
                GKNPVLTTTLNDDQLLKLSEQVQHDSEIIARLTDKKMKDLSEMSHTLTPENTLDISSL
                SSNAVSLIISVAVLLSALRTAETKLGSQLSLIAFDATKSAAENIVRQGLAALSSSITG
                AVTQVGITGIGAKKTHSGISDQKGALRKNLATAQSLEKELAGSKLGLNKQIDTNITSP
                QTNSSTKFLGKNKLAPDNISLSTEHKTSLSSPDISLQDKIDTQRRTYELNTLSAQQKQ
                NIGRATMETSAVAGNISTSGGRYASALEEEEQLISQASSKQAEEASQVSKEASQATNQ
                LIQKLLNIIDSINQSKNSAASQIAGNIRA"
     RBS        3971. .3977
                /gene="ipaD"
     gene       3971. .4761
                /gene="ipaD"
     CDS        3983. .>4761
                /gene="ipaD"
                /codon_start=1
                /transl_table=11
                /product="antigen"
                /protein_id="AAA98426.1"
                /db_xref="PID:g552045"
                /db_xref="GI:552045"
                /translation="MNITTLTNSISTSSFSPNNTNGSSTETVNSDIKTTTSSHPVSSL
                TMLNDTLHNIRTTNQALKKELSQKTLTKTSLEEIALHSSQISMDVNKSAQLLDILSRN
                EYPINKDARELLHSAPKEAELDGDQMISHRELWAKIANSINDINEQYLKVYEHAVSSY
                TQMYQDFSAVLSSLAGWISPGGNDGNSVKLQVNSLKKALEELKEKYKDKPLYPANNTV
                SQEQANKWLTELGGTIGKVSQKNGGYVVSINMTPIDNMLKS"
BASE COUNT     1769 a    952 c    805 g   1235 t
ORIGIN


  Query Match             6.3%;  Score 64.6;  DB 1;  Length 4761;
  Best Local Similarity   56.7%;  Pred. No. 3.9e-07;
  Matches  140;  Conservative    0;  Mismatches  104;  Indels    3;  Gaps    1;

Qy     381 AACCATTTCTGATGCTGAGATTTGGGATATGGTTTCCCAAAATATATCGGCGATAGGTGA 440
           ||  || ||| ||   ||  | |||| ||   || | |    ||    | ||   |||
Db    4363 AATGATATCTCATAGAGAACTGTGGGCTAAAATTGCAAACTCCATCAATGATATTAATGA 4422

Qy     441 CAGCTATCTGGGCGTTTATGAAAACGTTGTCGCAGTCTATACCGATTTTTATCAGGCCTT 500
           ||||||   || |||||| | |  ||       |||||  |  | ||||| |  ||
Db    4423 ACAGTATCTGAAAGTATATGAACATGCCGTTAGTTCATATACTCAAATGTATCAAGATTT 4482

Qy     501 CAGTGATATTCTTTCCAAAATGGGAGGCTGGTTATTACC---AGGTAAGGACGGTAATAC 557
            || | | |||||||||   | |  |||||| | |  ||   |||||| ||||| ||  |
Db    4483 TAGCGCTGTTCTTTCCAGTCTTGCCGGCTGGATCTCTCCCGGAGGTAACGACGGAAACTC 4542

Qy     558 CGTTAAGCTAGATGTTACCTCACTCAAAAATGATTTAAACAGTTTAGTCAATAAATATAA 617
           |||  ||  || | ||  | |||  || ||||| |  ||  |     |     | ||||||||
Db    4543 CGTGAAATTACAAGTCAACTCGCTTAAAAAGGCATTGGAAGAACTCAAGGAAAAATATAA 4602

Qy     618 TCAAATA 624
           | | |
Db    4603 AGATAAA 4609


RESULT   8
I66494/c
LOCUS       I66494       7218 bp    DNA             PAT       23-DEC-1997
DEFINITION  Sequence 14 from patent US 5670367.
ACCESSION   I66494
NID         g2724471
VERSION     I66494.1  GI:2724471
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 7218)
  AUTHORS   Dorner,F., Scheiflinger,F. and Falkner,F.Gunter.
  TITLE     Recombinant fowlpox virus
  JOURNAL   Patent: US 5670367-A 14 23-SEP-1997;
FEATURES             Location/Qualifiers
     source          1. .7218
                     /organism="unknown"
BASE COUNT     1944 a   1491 c   1486 g   1929 t    368 others
ORIGIN


  Query Match             4.7%;  Score 48.8;  DB 5;  Length 7218;
  Best Local Similarity   5.6%;  Pred. No. 0.0075;
```

```
     CDS             606. .1073
                     /note="18 kD protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA26521.1"
                     /db_xref="PID:g152741"
                     /db_xref="GI:152741"
                     /translation="MSLNITENESISTAVIDAINSGATLKDINAIPDDMMDDIYSYAY
                     DFYNKGRIEEAEVFFRFLCIYDFYNVDYIMGLAAIYQIKEQFQQAADLYAVAFALGKN
                     DYTPVFHTGQCQLRLKAPLKAKECFELVIQHSNDEKLKIKAQSYLDAIQDIKE"
     CDS             1079. .2821
                     /note="IpaB protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA26522.1"
                     /db_xref="PID:g152742"
                     /db_xref="GI:152742"
                     /translation="MHNVSTTTTGFPLAKILTSTELGDNTIQAANDAANKLFSLTIAD
                     LTANQNINTTNAHSTSNILIPELKAPKSLNASSQLTLLIGNLIQILGEKSLTALTNKI
                     TAWKSQQQARQQKNLEFSDKINTLLSETEGLTRDYEKQINKLKNADSKIKDLENKINQ
                     IQTRLSNLDPESPEKKKLSREEIQLTIKKDAAVKDRTLIEQKTLSIHSKLTDKSMQLE
                     KEIDSFSAFSNTASAEQLSTQQKSLTGLASVTQLMATFIQLVGKNNEESLKNDLALFQ
                     SLQESRKTEMERKSDEYAAEVRKAEELNRVMGCVGKILGALLTIVSVVAAAFSGGASL
                     ALAAVGLALMVTDAIVQAATGNSFMEQALNPIMKAVIEPLIKLLSDAFTKMLEGLGVD
                     SKKAKMIGSILGAIAGALVLVAAVVLVATVGKQAAAKLAENIGKIIGKTLTDLIPKFL
                     KNFSSQLDDLITNAVARLNKFLGAAGDEVISKQIISTHLNQAVLLGESVNSATQAGGS
                     VASAVFQNSASTNLADLTLSKYQVEQLSKYISEAIEKFGQLQEVIADLLASMSNSQAN
                     RTDVAKAILQQTTA"
     CDS             2784. .3932
                     /note="IpaC protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA26523.1"
                     /db_xref="PID:g152743"
                     /db_xref="GI:152743"
                     /translation="MLQKQFCNKLLLDTNKENVMEIQNTKPTQTLYTDISTKQTQSSS
                     ETQKSQNYQQIAAHIPLNVGKNPVLTTTLNDDQLLKLSEQVQHDSEIIARLTDKKMKD
                     LSEMSHTLTPENTLDISSLSSNAVSLIISVAVLLSALRTAETKLGSQLSLIAFDATKS
                     AAENIVRQGLAALSSSITGAVTQVGITGIGAKKTHSGISDQKGALRKNLATAQSLEKE
                     LAGSKLGLNKQIDTNITSPQTNSSTKFLGKNKLAPDNISLSTEHKTSLSSPDISLQDK
                     IDTQRRTYELNTLSAQQKQNIGRATMETSAVAGNISTSGGRYASALEEEEQLISQASS
                     KQAEEASQVSKEASQATNQLIQKLLNIIDSINQSKNSAASQIAGNIRA"
     CDS             3983. .4981
                     /note="IpaD protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA26524.1"
                     /db_xref="PID:g152744"
                     /db_xref="GI:152744"
                     /translation="MNITTLTNSISTSSFSPNNTNGSSTETVNSDIKTTTSSHPVSSL
                     TMLNDILHNIRTTNQALKKELSQKTLTKTSLEEIALHSSQISMDVNKSAQLLDILSRN
                     EYPINKDARELLHSAPKEAELDGDQMISHRELWAKIANSINDINEQYLKVYEHAVSSY
                     TQMYQDFSAVLSSLAGWISPGGNDGNSVKLQVNSLKKALEELKEKYKDKPLYPANNTV
                     SQEQANKWLTELGGTIGKVSQKNGGYVVSINMTPIDNMLKSLDNLGGNGEVVLDNAKY
                     QAWNAGFSAEDETMKNNLQTLVQKYSNANSIFDNLVKVLSSTISSCTDTDKLFLHF"
     CDS             4990. .>5255
                     /note="IpaA protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA26525.1"
                     /db_xref="PID:g551981"
                     /db_xref="GI:551981"
                     /translation="MHNVNNTQAPTFLYKATSPSSTEYSELKSKISDIHSSQTSLKTP
                     ASVSEKENFATSFNQKCLDFLFSSSGKEDVLRSIYSNSMNAYAK"
BASE COUNT      1950 a    1037 c     889 g    1379 t
ORIGIN


  Query Match             9.8%;  Score 101.6;  DB 1;  Length 5255;
  Best Local Similarity   51.7%;  Pred. No. 4.2e-17;
  Matches  335;  Conservative    0;  Mismatches  274;  Indels   39;  Gaps    3;

Qy     381 AACCATTTCTGATGCTGAGATTTGGGATATGGTTTCCCAAAATATATCGGCGATAGGTGA 440
```

```
Db    4363 AATGATATCTCATAGAGAACTGTGGGCTAAAATTGCAAACTCCATCAATGATATTAATGA 4422

Qy     441 CAGCTATCTGGGCGTTTATGAAAACGTTGTCGCAGTCTATACCGATTTTTATCAGGCCTT 500
           ||||||   || ||||||  | |  ||       |||||  |  | |||||  |  ||
Db    4423 ACAGTATCTGAAAGTATATGAACATGCCGTTAGTTCATATACTCAAATGTATCAAGATTT 4482

Qy     501 CAGTGATATTCTTTCCAAAATGGGAGGCTGGTTATTACC---AGGTAAGGACGGTAATAC 557
           ||  |  |||||||||    | |  ||||||  |  |  ||   ||||||  ||||  |
Db    4483 TAGCGCTGTTCTTTCCAGTCTTGCCGGCTGGATCTCTCCCGGAGGTAACGACGGAAACTC 4542

Qy     558 CGTTAAGCTAGATGTTACCTCACTCAAAAATGATTTAAACAGTTTAGTCAATAAATATAA 617
           |||  ||  ||  |  ||  |  |||  ||  |||||  |  ||  |     |  |||||||
Db    4543 CGTGAAATTACAAGTCAACTCGCTTAAAAAGGCATTGGAAGAACTCAAGGAAAAATATAA 4602

Qy     618 TCAAATAAACAGTAATACCGTTTTATTTCCAGCGCAGTCAGGCAGCGGCGTTAAAGTAGC 677
           |  |  |     ||  |  |                                |||    |
Db    4603 AGATAAACCGCTATATCCAGCAAA--------------------------TAATACTGT 4635

Qy     678 CACTGAAGCGGAAGCGAGACAGTGGCTCAGTGAATTGAATTTACCGAATAGCTGCCTGAA 737
           |  |  |   ||||  |   |  |||||  |  |||||   |  |  |  |   ||    |
Db    4636 TAGTCAGGAACAAGCAAATAAATGGCTTACAGAATTAGGTGGAACAATCGGCAAGGTATC 4695

Qy     738 ATCTTATGGATCCGGTTATGTCGTCACCGTTGATCTGACGCCATTACAAAAAATGGTTCA 797
           |       ||  |||||  ||||   |  |  ||||  |||  ||  |  ||  |||  |  |
Db    4696 TCAAAAAAACGGGGGATATGTTGTCAGTATAAACATGACCCCAATAGACAATATGTTAAA 4755

Qy     798 GGATATTGATGGTTTAGGCGCGCCGGGAAAAGACTCAAAACTCGAAATGGATAACGCCAA 857
           |  |||  | ||||  |     ||  |  |         |  |   | |||||  ||  ||
Db    4756 AAGCTTAGATAATCTAGGTGGAAATGGCGAGG---------TTGTGCTAGATAATGCAAA 4806

Qy     858 ATATCAAGCCTGGCAGTCGGGTTTTAAAGCGCAGGAAGAAAATATGAAAACCACATTACA 917
           ||||||  ||  |||  |   |  ||  ||     ||   |  ||  ||||  |||||||  |   |  ||
Db    4807 ATATCAGGCATGGAATGCCGGATTCTCTGCCGAAGATGAAACAATGAAAAATAATCTTCA 4866

Qy     918 GACGCTGACGCAAAAATATAGCAATGCCAATTCATTGTACGACAACCTGGTAAAAGTGCT 977
           ||  |    |||||||| ||  |||||||||     |  |  ||  ||  |  |||||  ||  |
Db    4867 AACTTTAGTTCAAAAATACAGTAATGCCAATAGTATTTTTGATAATTTAGTAAAGGTTTT 4926

Qy     978 GAGCAGTACGATAAGTAGCAGCCTGGAAACCGCCAAAAGCTTCCTGCA 1025
           |||  |||||  |||||      |    ||  ||  |   |||    ||  ||  ||
Db    4927 GAGTAGTACAATAAGCTCATGTACAGATACAGATAAACTTTTTCTCCA 4974
```

```
RESULT   7
W10PPIBCA
LOCUS       W10PPIBCA     4761 bp    DNA             BCT       07-MAY-1996
DEFINITION  Plasmid pWR100 (from S.flexneri) antigen ippI, ipaB, ipaC genes,
            complete cds, and ippD gene, 3' end.
ACCESSION   M34849
NID         g155327
VERSION     M34849.1  GI:155327
KEYWORDS    antigen.
SOURCE      Plasmid pWR100 (clone: HB101/pHS[4108,5357].).
  ORGANISM  Plasmid pWR100
            plasmids.
REFERENCE   1  (bases 1 to 4761; 1 to 538; 3987 to 4761)
  AUTHORS   Baudry,B., Kaczorek,M. and Sansonetti,P.J.
  TITLE     Nucleotide sequence of the invasion plasmid antigen B and C genes
            (ipaB and ipaC) of Shigella flexneri
  JOURNAL   Microb. Pathog. 4 (5), 345-357 (1988)
  MEDLINE   89200844
COMMENT     Draft entry and computer-readable sequence for [1] kindly submitted
            by B.Baudry, 05-JUN-1990.
               Center for Vaccine Development
               University of Maryland at Baltimore
               10 South Pine Street
               Baltimore, MD 21201.
FEATURES             Location/Qualifiers
     source          1. .4761
                     /organism="Plasmid pWR100"
                     /plasmid="Plasmid pWR100"
                     /specific_host="M90T"
```

```
     CDS             2053. .2520
                     /note="18k gene (AA 1-155)"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA33380.1"
                     /db_xref="PID:g47054"
                     /db_xref="GI:47054"
                     /db_xref="SWISS-PROT:P18008"
                     /translation="MSLNITENESISTAVIDAINSGATLKDINAIPDDMMDDIYSYAY
                     DFYNKGRIEEAEVFFRFLCIYDFYNVDYIMGLAAIYQIKEQFQQAADLYAVAFALGKN
                     DYTPVFHTGQCQLRLKAPLKAKECFELVIQHSNDEKLKIKAQSYLDAIQDIKE"
     RBS             2513. .2515
                     /note="put. ribosome binding site"
     CDS             2526. .4268
                     /note="ipaB gene (AA 1-580)"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA33381.1"
                     /db_xref="PID:g47055"
                     /db_xref="GI:47055"
                     /db_xref="SWISS-PROT:P18011"
                     /translation="MHNVSTTTTGFPLAKILASTELGDNTIQAANDAANKLFSLTIAD
                     LTANQNINTTNAHSTSNILIPELKAPKSLNASSQLTLLIGNLIQILGEKSLTALTNKI
                     TAWKSQQQARQQKNLEFSDKINTLLSETEGLTRDYEKQINKLKNADSKIKDLENKINQ
                     IQTRLSELDPESPEKKKLSREEIQLTIKKDAAVKDRTLIEQKTLSIHSKLTDKSMQLE
                     KEIDSFSAFSNTASAEQLSTQQKSLTGLASVTQLMATFIQLVGKNNEESLKNDLALFQ
                     SLQESRKTEMERKSDEYAAEVRKAEELNRVMGCVGKILGALLTIVSVVAAAFSGGASL
                     ALAAVGLALMVTDAIVQAATGNSFMEQALNPIMKAVIEPLIKLLSDAFTKMLEGLGVD
                     SKKAKMIGSILGAIAGALVLVAAVVLVATVGKQAAAKLAENIGKIIGKTLTDLIPKFL
                     KNFSSQLDDLITNAVARLNKFLGAAGDEVISKQIISTHLNQAVLLGESVNSATQAGGS
                     VASAVFQNSASTNLADLTLSKYQVEQLSKYISEAIEKFGQLQEVIADLLASMSNSQAN
                     RTDVAKAILQQTTA"
     RBS             4224. .4226
                     /note="put. ribosome binding site"
     CDS             4231. .5379
                     /note="ipaC gene (AA 1-382)"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA33382.1"
                     /db_xref="PID:g47056"
                     /db_xref="GI:47056"
                     /db_xref="SWISS-PROT:P18012"
                     /translation="MLQKQFCNKLLLDTNKENVMEIQNTKPTQILYTDISTKQTQSSS
                     ETQKSQNYQQIAAHIPLNVGKNPVLTTTLNDDQLLKLSEQVQHDSEIIARLTDKKMKD
                     LSEMSHTLTPENTLDISSLSSNAVSLIISVAVLLSALRTAETKLGSQLSLIAFDATKS
                     AAENIVRQGLAALSSSITGAVTQVGITGIGAKKTHSGISDQKGALRKNLATAQSLEKE
                     LAGSKLGLNKQIDTNITSPQTNSSTKFLGKNKLAPDNISLSTEHKTSLSSPDISLQDK
                     IDTQRRTYELNTLSAQQKQNIGRATMETSAVAGNISTSGGRYASALEEEEQLISQASS
                     KQAEEASQVSKEASQATNQLIQKLLNIIDSINQSKNSTASQIAGNIRA"
     promoter        5376. .5381
                     /note="-35 region for 4.5kb transcript"
     promoter        5396. .5401
                     /note="-10 region for 4.5kb transcript"
     RBS             5421. .5423
                     /note="put. ribosome binding site"
     CDS             5430. .6428
                     /note="ipaD gene (AA 1-332)"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA33383.1"
                     /db_xref="PID:g47057"
                     /db_xref="GI:47057"
                     /db_xref="SWISS-PROT:P18013"
                     /translation="MNITTLTNSISTSSFSPNNTNGSSTETVNSDIKTTTSSHPVSSL
                     TMLNDTLHNIRTTNQALKKELSQKTLTKTSLEEIALHSSQISMDVNKSAQLLDILSRH
                     EYPINKDARELLHSAPKEAELDGDQMISHRELWAKIANSINDINEQYLKVYEHAVSSY
                     TQMYQDFSAVLSSLAGWISPGGNDGNSVKLQVNSLKKALEELKEKYKDKPLYPANNTV
                     SQEQANKWLTELGGTIGKVSQKNGGYVVSINMTPIDNMLKSLDNLGGNGEVVLDNAKY
                     QAWNAGFSAEDETMKNNLQTLVQKYSNANSIFDNLVKVLSSTISSCTDTDKLFLHF"
BASE COUNT      2845 a   1501 c   1246 g   2071 t
ORIGIN
```

```
  Query Match             9.8%;  Score 101.6;  DB 1;  Length 7663;
  Best Local Similarity   51.7%;  Pred. No. 4.4e-17;
  Matches  335;  Conservative    0;  Mismatches  274;  Indels   39;  Gaps    3;

Qy    381 AACCATTTCTGATGCTGAGATTTGGGATATGGTTTCCCAAAATATATCGGCGATAGGTGA 440
          || || ||| ||   ||  | |||| ||   || |  |    ||    |  ||   |||
Db   5810 AATGATATCTCATAGAGAACTGTGGGCTAAAATTGCAAACTCCATCAATGATATTAATGA 5869

Qy    441 CAGCTATCTGGGCGTTTATGAAAACGTTGTCGCAGTCTATACCGATTTTTATCAGGCCTT 500
          ||||||   || ||||||  | |  ||       |||||  |  | |||||  |  ||
Db   5870 ACAGTATCTGAAAGTATATGAACATGCCGTTAGTTCATATACTCAAATGTATCAAGATTT 5929

Qy    501 CAGTGATATTCTTTCCAAAATGGGAGGCTGGTTATTACC---AGGTAAGGACGGTAATAC 557
          || | | ||||||||||   | |  ||||||  | |  ||   ||||||  |||||  ||  |
Db   5930 TAGCGCTGTTCTTTCCAGTCTTGCCGGCTGGATCTCTCCCGGAGGTAACGACGGAAACTC 5989

Qy    558 CGTTAAGCTAGATGTTACCTCACTCAAAAATGATTTAAACAGTTTAGTCAATAAATATAA 617
          |||  ||  || |  ||  | |||  ||  |||||  |  ||  |      |      | |||||||||
Db   5990 CGTGAAATTACAAGTCAACTCGCTTAAAAAGGCATTGGAAGAACTCAAGGAAAAATATAA 6049

Qy    618 TCAAATAAACAGTAATACCGTTTTATTTCCAGCGCAGTCAGGCAGCGGCGTTAAAGTAGC 677
          | | |        || | |                                   |||    |
Db   6050 AGATAAACCGCTATATCCAGCAAA----------------------------TAATACTGT 6082

Qy    678 CACTGAAGCGGAAGCGAGACAGTGGCTCAGTGAATTGAATTTACCGAATAGCTGCCTGAA 737
          | | | |   ||||  |   | |||||  |  |||||   |  | | |   ||     |
Db   6083 TAGTCAGGAACAAGCAAATAAATGGCTTACAGAATTAGGTGGAACAATCGGCAAGGTATC 6142

Qy    738 ATCTTATGGATCCGGTTATGTCGTCACCGTTGATCTGACGCCATTACAAAAAATGGTTCA 797
                |       || ||||| ||||   |  |  ||||  ||| || | ||  ||| |  |
Db   6143 TCAAAAAAACGGGGGATATGTTGTCAGTATAAACATGACCCCAATAGACAATATGTTAAA 6202

Qy    798 GGATATTGATGGTTTAGGCGCGCCGGGAAAAGACTCAAAACTCGAAATGGATAACGCCAA 857
                | |||  | ||||  |      ||  | |          | |   | |||||  || ||
Db   6203 AAGCTTAGATAATCTAGGTGGAAATGGCGAGG---------TTGTGCTAGATAATGCAAA 6253

Qy    858 ATATCAAGCCTGGCAGTCGGGTTTTAAAGCGCAGGAAGAAAATATGAAAACCACATTACA 917
          ||||||  || |||  |  | || ||    ||  |  || ||||  |||||||  |   | ||
Db   6254 ATATCAGGCATGGAATGCCGGATTCTCTGCCGAAGATGAAACAATGAAAAATAATCTTCA 6313

Qy    918 GACGCTGACGCAAAAATATAGCAATGCCAATTCATTGTACGACAACCTGGTAAAAGTGCT 977
            ||  |    |||||||| || |||||||||    | |  || ||  | |||||  ||  |
Db   6314 AACTTTAGTTCAAAAATACAGTAATGCCAATAGTATTTTTGATAATTTAGTAAAGGTTTT 6373

Qy    978 GAGCAGTACGATAAGTAGCAGCCTGGAAACCGCCAAAAGCTTCCTGCA 1025
          |||  |||||  ||||      |    || || |  |||   || || ||
Db   6374 GAGTAGTACAATAAGCTCATGTACAGATACAGATAAACTTTTTCTCCA 6421
```

```
RESULT   6
SHFIPABCD
LOCUS       SHFIPABCD    5255 bp    DNA             BCT       19-MAR-1991
DEFINITION  S.flexneri ipaBCD region encoding invasion plasmid antigens.
ACCESSION   J04117
NID         g152740
VERSION     J04117.1  GI:152740
KEYWORDS    ipaA gene; ipaB gene; ipaC gene; ipaD; membrane protein.
SOURCE      S.flexeri (strain M90T-W) DNA, clone pHC17.
  ORGANISM  Shigella flexneri
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Shigella.
REFERENCE   1  (bases 1 to 5255)
  AUTHORS   Venkatesan,M.M., Buysse,J.M. and Kopecko,D.J.
  TITLE     Characterization of invasion plasmid antigen genes (ipaBCD) from
            Shigella flexneri
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 85, 9317-9321 (1988)
  MEDLINE   89057927
COMMENT     Draft entry and clean copy of sequence [1] kindly submitted by
            M.M.Venkatesan, 03-JAN-1989.
FEATURES             Location/Qualifiers
     source          1. .5255
                     /organism="Shigella flexneri"
                     /db_xref="taxon:623"
```

```
                     LTANKNINTTNAHSTSNILIPELKAPKSLNASSQLTLLIGNLIQILGEKSLTALTNKI
                     IAWKSQQQARQQKNLEFSDKINTLLSETEGLTRDYEKQINKLKNADSKIKDLENKINQ
                     IQTRLSELDPDSPEKKKLSREEIQLTIKKDAAVKDRTLIEQKTLSIHSKLTDKSMQLE
                     KEIDSFSAFSNTASAEQLSTQQKSLTGLASVTQLMATFIQLVGKNNEESLKNDLALFQ
                     SLQESRKTEMERKSDEYAAEVRKAEELNRVMGCVGKILGALLTIVSVVAAAFSGGASL
                     ALADVGLALMVTDAIVQAATGNSFMEQALNPIMKAVIEPLIKLLSDAFTKMLEGLGVD
                     SKKAKMIGSILGAIAGALVLVAAVVLVATVGKQAAAKLAENIGKIIGKTLTDLIPKFL
                     KNFSSQLDDLIITNAVARLNKFLGAAGDEVISKQIISTHLNQAVLLGESVNSATQAGGS
                     VASAVFQNSASTNLADLTLSKYQVEQLSKYISEAIEKFGQLQEVIADLLASMSNSQAN
                     RTDVAKAILQQTTA"
     CDS             3178. .4326
                     /note="IpaC ORF"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA43191.1"
                     /db_xref="PID:g46936"
                     /db_xref="GI:46936"
                     /db_xref="SWISS-PROT:Q03946"
                     /translation="MLQKQFCNKLLLDTNKENVMEIQNTKSTQILYTDISTKQTQSSS
                     ETQKSQNYQQIAAHIPLNVGKNPVLTTTLNDDQLLKLSEQVQHDSEIIARLTDKKMKD
                     LSEMSHTLTPENTLDISSLSSNAVSLIISVAVLLSALRTAETKLGSQLSLIAFDATKS
                     AAENIVRQGLAALSSSITGAVTQVGITGIGAKKTHSGISDQKGALRKNLATAQSLEKE
                     LAGSKLGLNKQIDTNITSPQTNSSTKFLGKNKLAPDNISLSTEHKTSLSSPDISLQDK
                     IDTQRRTYELNTLSAQQKQNIGRATMETSAVAGNISTSGGRYASALEEEEQLISQASS
                     KQAEEASQVSKEASQATNQLIQKLLNIIDNINQSRSSTASQIAGNIRA"
     CDS             4377. .5375
                     /note="IpaD ORF"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA43192.1"
                     /db_xref="PID:g46937"
                     /db_xref="GI:46937"
                     /db_xref="SWISS-PROT:Q03947"
                     /translation="MNITTLTNSISTSSFSPNNTNGSSTETVNSDIKTTTSSHPVSSL
                     TMLNDTLHNIRTTNQALKKELSQKTLTKTSLEEIALHSSQISMDVNKSAQLLNILSKT
                     EYPINKDARELLHSAPKEAELDGYEMISHRELWAKIANSINDINEQYLKVYEHAVSSY
                     TQMYQEFSAVLSSLAGWISPGGNDGNSVKLQVKSLKDALTTLKKNYEDKPLYPATNTV
                     SEQEANKWLTELGGTIGTVSAKNGGYVVSINMTPIYNMLNRLDNLGGNGEVVLDNAKY
                     QAWNAGFSAEDEIMKNNLQTLVQKYSNANSIFDNLVKVLSSTISSCTDTDKLFLHF"
BASE COUNT     2518 a   1369 c   1135 g   1847 t
ORIGIN


  Query Match             9.8%;  Score 101.6;  DB 1;  Length 6869;
  Best Local Similarity   51.7%;  Pred. No. 4.3e-17;
  Matches  335;  Conservative    0;  Mismatches  274;  Indels   39;  Gaps    3;

Qy     381 AACCATTTCTGATGCTGAGATTTGGGATATGGTTTCCCAAAATATATCGGCGATAGGTGA 440
           ||  || ||| ||   ||  | |||| ||   || |  |    ||    |  ||   |||
Db    4757 AATGATATCTCATAGAGAACTGTGGGCTAAAATTGCAAACTCCATCAATGATATTAATGA 4816

Qy     441 CAGCTATCTGGGCGTTTATGAAAACGTTGTCGCAGTCTATACCGATTTTTATCAGGCCTT 500
           ||||||   || |||||| | |  ||       |||||  |  | ||||| |  ||
Db    4817 ACAGTATCTGAAAGTATATGAACATGCCGTTAGTTCATATACTCAAATGTATCAAGAATT 4876

Qy     501 CAGTGATATTCTTTCCAAAATGGGAGGCTGGTTATTACC---AGGTAAGGACGGTAATAC 557
            || | | |||||||||    | |  |||||| | |  ||   |||||| ||||| ||  |
Db    4877 TAGCGCTGTTCTTTCCAGTCTTGCCGGCTGGATCTCTCCCGGAGGTAACGACGGAAACTC 4936

Qy     558 CGTTAAGCTAGATGTTACCTCACTCAAAAATGATTTAAACAGTTTAGTCAATAAATATAA 617
           |||  ||  || | || |  || || ||| | |  || |  |    |    || || ||| |
Db    4937 CGTGAAATTACAAGTCAAGTCGCTTAAAGACGCATTGACAACACTCAAGAAAAACTATGA 4996

Qy     618 TCAAATAAACAGTAATACCGTTTTATTTCCAGCGCAGTCAGGCAGCGGCGTTAAAGTAGC 677
           | | |       || | |                                 |||    |
Db    4997 AGATAAACCGCTATATCCAGCAAC--------------------------TAATACTGT 5029

Qy     678 CACTGAAGCGGAAGCGAGACAGTGGCTCAGTGAATTGAATTTACCGAATAGCTGCCTGAA 737
           | |||    |||| |   | ||||| |  |||||   |  | | |   ||    |
Db    5030 TAGTGAGCAAGAAGCAAATAAATGGCTTACAGAATTAGGTGGAACAATCGGCACGGTATC 5089

Qy     738 ATCTTATGGATCCGGTTATGTCGTCACCGTTGATCTGACGCCATTACAAAAAATGGTTCA 797
           |  |       || ||||| ||||  |  |  |||| ||| || | || ||| |  |
Db    5090 TGCAAAAAACGGGGGATATGTTGTCAGTATAAACATGACCCCAATATATAATATGTTAAA 5149

Qy     798 GGATATTGATGGTTTAGGCGCGCCGGGAAAAGACTCAAAACTCGAAATGGATAACGCCAA 857
           |  |||  | |||| |      ||  | |          | |   | ||||| || ||
Db    5150 CCGCTTAGATAATCTAGGTGGAAATGGCGAGG---------TTGTGCTAGATAATGCAAA 5200

Qy     858 ATATCAAGCCTGGCAGTCGGGTTTTAAAGCGCAGGAAGAAAATATGAAAACCACATTACA 917
           |||||| || ||| |  | || ||    ||  | || ||||  |||||||  |   | ||
Db    5201 ATATCAGGCATGGAATGCCGGATTCTCTGCCGAAGATGAAACAATGAAAAATAATCTTCA 5260

Qy     918 GACGCTGACGCAAAAATATAGCAATGCCAATTCATTGTACGACAACCTGGTAAAAGTGCT 977
           ||  |    |||||||| || ||||||||||    | |  || ||  | ||||| ||  |
Db    5261 AACTTTAGTTCAAAAATACAGTAATGCCAATAGTATTTTTGATAATTTAGTAAAGGTTTT 5320

Qy     978 GAGCAGTACGATAAGTAGCAGCCTGGAAACCGCCAAAAGCTTCCTGCA 1025
           ||| ||||| |||||      |    || || |  |||   || || ||
Db    5321 GAGTAGTACAATAAGCTCATGTACAGATACAGATAAACTTTTTCTCCA 5368


RESULT   5
SFPVR2
LOCUS       SFPVR2       7663 bp    DNA             BCI       12-SEP-1993
DEFINITION  Shigella flexineri plasmid YSH6000 virulence region-2 DNA
            containing 24k gene, 18k gene, ipaB gene, ipaC gene and ipaD gene.
ACCESSION   X15319
NID         g47052
VERSION     X15319.1  GI:47052
KEYWORDS    ipaB gene; ipaC gene; ipaD gene; virulence proteins.
SOURCE      Enterobacteriaceae.
  ORGANISM  Enterobacteriaceae
            Bacteria; Proteobacteria; gamma subdivision.
REFERENCE   1  (bases 1 to 7663)
  AUTHORS   Sasakawa,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAY-1989) Sasakawa C., Institute of Medical Science,
            University of Tokyo, Shirokanedai 4-6-1, Minato-ku Tokyo 108, Japan
REFERENCE   2  (bases 1010 to 6538)
  AUTHORS   Sasakawa,C., Adler,B., Tobe,T., Okada,N., Nagai,S., Komatsu,K. and
            Yoshikawa,M.
  TITLE     Functional organization and nucleotide sequence of virulence
            Region-2 on the large virulence plasmid in Shigella flexneri 2a
  JOURNAL   Mol. Microbiol. 3 (9), 1191-1201 (1989)
  MEDLINE   90014179
COMMENT     Region 2 has three promoters that transcribe RNAs of 4.0, 4.5 and
            7.5kb. The 4.0 kb transcript produces the 24,18kD, IpaB and IpaC
            proteins. The 4.5 kb transcript produces the ipaD protein. The
            7.5kb transcript covers the entire region 2.
            Data kindly reviewed (20-NOV-1989) by Sasakawa C.
FEATURES             Location/Qualifiers
     source          1. .7663
                     /organism="Enterobacteriaceae"
                     /plasmid="YSH6000"
                     /db_xref="taxon:543"
     promoter        1291. .1296
                     /note="-35 region for 4kb transcript"
     promoter        1312. .1318
                     /note="-10 region for 4kb transcript"
     RBS             1360. .1362
                     /note="put. ribosome binding site"
     CDS             1370. .1996
                     /note="24k gene (AA 1-208)"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA33379.1"
                     /db_xref="PID:g47053"
                     /db_xref="GI:47053"
                     /db_xref="SWISS-PROT:P33548"
                     /translation="MQILNKILPQVEFAIPRPSFDSLSRNKLVKKILSVFNLKQRFPQ
                     KNFGCPVNINKIRDSVIDKIKDSNSGNQLFCWMSQERTTYVSSMINRSIDEMAIHNGV
                     VLTSDNKKNIFAAIEKKFPDIKLDEKSAQTSISHTALNEIASSGLRAKILKRYSSDMD
                     LFNTQMKDLTNLVSSSVYDKIFNESTKVLQIEISAEVLKAVYRQSNTN"
     RBS             2043. .2045
                     /note="put. ribosome binding site"
```

```
                     /translation="MVTSVRTQPPVIMPGMQTEIKTQATNLAANLSAVRESATATLSG
                     EIKGQQLEDFPALIKQASLDALFKCGKDAEALKEVFTNSNNVAGKKAIMEFAGLFRSA
                     LNATSDSPEAKTLLMKVGAEYTAQIIKDGLKEKSAFGPWLPETKKAEAKLENLEKQLL
                     DIIKNNTGGELSKLSTNLVMQEVMPYIASCIEHNFGCTLDPLTRSSLTQLVDKAAAKA
                     VEALDMCHQKLTQEQGT"
BASE COUNT     1584 a   1156 c   1458 g   1195 t
ORIGIN


  Query Match           79.3%;  Score 818;  DB 1;  Length 5393;
  Best Local Similarity  89.1%;  Pred. No. 2.7e-210;
  Matches  920;  Conservative    0;  Mismatches  100;  Indels   12;  Gaps    3;

Qy      1 ATGCTTAATATTCAAAATTATTCCGCTTCTCCTCATCCGGGGATCGTTGCCGAACGGCCG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3651 ATGCTTAATATTCAAAATTATTCCGCTTCTCCTCATCCGGGGATCGTTGCCGAACGGCCG 3710

Qy     61 CAGACTCCCTCGGCGAGCGAGCACGTCGAGACTGCCGTGGTACCGTCTACCACAGAACAT 120
          |||||||| |||||||||||||||| ||||| ||||||||||||||||||||||||||||
Db   3711 CAGACTCCTTCGGCGAGCGAGCACGCCGAGATTGCCGTGGTACCGTCTACCACAGAACAT 3770

Qy    121 CGCGGTACAGATATCATTTCATTATCGCAGGCGGCTACTAAAATCCACCAGGCACAGCAG 180
          ||||| |||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db   3771 CGCGGCACAGATATCATTTCATTATCGCAGGCGGCTACTAAAATCCAGCAGGCACAGCAG 3830

Qy    181 ACGCTGCAGTCAACGCCACCGATCTCTGAAGAGAATAATGACGAGCGCACGCTGGCGCGC 240
          |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db   3831 ACGCTGCAGTCAACGCCACCGATTTCTGAAGAGAATAATGACGAGCGCACGCTGGCGCGC 3890

Qy    241 CAGCAGTTGACCAGCAGCCTGAATGCGCTGGCGAAGTCCGGCGTGTCATTATCCGCAGAA 300
          || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3891 CAACAGTTGACCAGCAGCCTGAATGCGCTGGCGAAGTCCGGCGTGTCATTATCCGCAGAA 3950

Qy    301 CAAAATGAGAACCTGCGGAGCGCGTTTTCTGCGCCGACGTCGGCCTTATTTAGCGCTTCG 360
          |||||||||||||||||||| |||||||||| ||||||||||||||||||||||||||||
Db   3951 CAAAATGAGAACCTGCGGAGCACGTTTTCTGCG-CGACGTCGGCCTTATTTAGCGCTTCG 4009

Qy    361 CCTATGGCGCAGCCGAGAACAACCATTTCTGATGCTGAGATTTGGGATATGGTTTCCCAA 420
          ||||||  ||   ||||||||||||||||||||||||||||||||||||||||||||||
Db   4010 CCTATG--GCCAGCGAGAACAACCATTTCTGATGCTGAGATTTGGGATATGGTTTCCCAA 4067

Qy    421 AATATATCGGCGATAGGTGACAGCTATCTGGGCGTTTATGAAAACGTTGTCGCAGTCTAT 480
          |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db   4068 AATATATCGGCGATAGGTGACAGCTACCTGGGCGTTTATGAAAACGTTGTCGCAGTCTAT 4127

Qy    481 ACCGATTTTTATCAGGCCTTCAGTGATATTCTTTCCAAAATGGGAGGCTGGTTATTACCA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  ||
Db   4128 ACCGATTTTTATCAGGCCTTCAGTGATATTCTTTCCAAAATGGGAGGCTGGTTATCGCCT 4187

Qy    541 GGTAAGGACGGTAATACCGTTAAGCTAGATGTTACCTCACTCAAAAATGATTTAAACAGT 600
          |||||||| || |||||| |||||||| |||||  |||||| |||| ||| ||| |||
Db   4188 GGTAAGGATGGAAATACCATTAAGCTAAATGTTGACTCACTTAAAAGTGAAATAAGTAGT 4247

Qy    601 TTAGTCAATAAATATAATCAAATAAACAGTAATACCGTTTTATTTCCAGCGCAGTCAGGC 660
          ||| | |||||||| | ||||||||| |  |||||  |||||||||  ||||  | |||
Db   4248 TTAATTAATAAATACACTCAAATAAATAAAAATACGATTTTATTTCCCTCGCAAACTGGC 4307

Qy    661 AGCGGCGTTAAAGTAGCCACTGAAGCGGAAGCGAGACAGTGGCTCAGTGAATTGAATTTA 720
          |||||  | | |  ||| ||  ||||||||||     |||||| | |  ||||||||||||
Db   4308 AGCGGAATGACAACAGCAACGAAAGCGGAAGCTGAGCAGTGGATTAAAGAATTGAATTTA 4367

Qy    721 CCGAATAGCTGCCTGAAATCTTATGGATCCGGTTATGTCGTCACCGTTGATCTGACGCCA 780
          ||| | ||||| || ||  | | ||| || ||||||||||    || |||  |  ||||
Db   4368 CCGGACAGCTGTCTAAAGGCGTCTGGTTCTGGTTATGTCGTACTGGTGGATACGGGGCCA 4427

Qy    781 TTACAAAAAATGGTTCAGGATATTGATGGTTTAGGCGCGCCCGGGAAAAGACTCAAAACTC 840
            |  |||||||||   ||| || ||||  ||||  ||          |  |||   ||
Db   4428 CTGAGCAAAATGGTTAGCGATCTTAATGGAATAGGATCG---------GGTTCAGCCCTT 4478

Qy    841 GAAATGGATAACGCCAAATATCAAGCCTGGCAGTCGGGTTTTAAAGCGCAGGAAGAAAAT 900
          ||| ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db   4479 GAACTGGATAACGCCAAATATCAAGCCTGGCAGTCGGGTTTTAAAGCACAGGAAGAAAAT 4538

Qy    901 ATGAAAACCACATTACAGACGCTGACGCAAAAATATAGCAATGCCAATTCATTGTACGAC 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4539 CTGAAAACCACATTACAGACGCTGACGCAAAAATATAGCAATGCCAATTCATTGTACGAC 4598

Qy    961 AACCTGGTAAAAGTGCTGAGCAGTACGATAAGTAGCAGCCTGGAAACCGCCAAAAGCTTC 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4599 AACCTGGTAAAAGTGCTGAGCAGTACGATAAGTAGCAGCCTGGAAACCGCCAAAAGCTTC 4658

Qy   1021 CTGCAAGGATAA 1032
          ||||||||||||
Db   4659 CTGCAAGGATAA 4670


RESULT   4
SDIPABCD
LOCUS       SDIPABCD     6869 bp    DNA             BCT       24-JUL-1991
DEFINITION  Shigella dysenteriae ipaBCD genes.
ACCESSION   X60777
NID         g46932
VERSION     X60777.1  GI:46932
KEYWORDS    ipaB gene; ipaC gene; ipaD gene; ippI gene.
SOURCE      Shigella dysenteriae.
  ORGANISM  Shigella dysenteriae
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Shigella.
REFERENCE   1  (bases 1 to 6869)
  AUTHORS   Palchaudhuri,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUN-1991) S. Palchaudhuri, Wayne State Univers.School
            of Medicine, Department of Immunology and, Microbiology, 540 E.
            Canfield, Detroit, MI 48201, USA
REFERENCE   2  (bases 1 to 6869)
  AUTHORS   Yao,R. and Palchaudhuri,S.
  TITLE     Nucleotide sequence of the ipaBCD structure genes of Shigella
            dysenteriae
  JOURNAL   Unpublished
FEATURES             Location/Qualifiers
     source          1. .6869
                     /organism="Shigella dysenteriae"
                     /strain="CG097"
                     /db_xref="taxon:622"
     CDS             317. .943
                     /note="24kD ORF"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA43188.1"
                     /db_xref="PID:g46933"
                     /db_xref="GI:46933"
                     /db_xref="SWISS-PROT:Q04314"
                     /translation="MQILNKILPQVEFAIPRPSFNSLSRNKLVKKILSVFNLKQRFPQ
                     KNFGCTVNINKIRDNVIDKIKDSNSGNQLFCWMSQERTTYVSSMINRSIDEMAIHNGV
                     VLTSDNKKNIFAAIEKKFPDIKLDEKSAQTSISHTALNEIASSGLRAKILKRYSSDMD
                     LFNTQMKDLTNLVSSSVYDKIFNESTKVLQIEISAEVLKAVYRQSNTN"
     CDS             1000. .1467
                     /note="IppI ORF"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA43189.1"
                     /db_xref="PID:g46934"
                     /db_xref="GI:46934"
                     /db_xref="SWISS-PROT:P18008"
                     /translation="MSLNITENESISTAVIDAINSGATLKDINAIPDDMMDDIYSYAY
                     DFYNKGRIEEAEVFFRFLCIYDFYNVDYIMGLAAIYQIKEQFQQAADLYAVAFALGKN
                     DYTPVFHTGQCQLRLKAPLKAKECFELVIQHSNDEKLKIKAQSYLDAIQDIKE"
     CDS             1473. .3215
                     /note="IpaB ORF"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA43190.1"
                     /db_xref="PID:g46935"
                     /db_xref="GI:46935"
                     /db_xref="SWISS-PROT:Q03945"
                     /translation="MHNVNTTTTGLSLAKILASTELGDNTIQAGNDAANKLFSLTIAD
```

```
Qy    541 GGTAAGGACGGTAATACCGTTAAGCTAGATGTTACCTCACTCAAAAATGATTTAAACAGT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    624 GGTAAGGACGGTAATACCGTTAAGCTAGATGTTACCTCACTCAAAAATGATTTAAACAGT 683

Qy    601 TTAGTCAATAAATATAATCAAATAAACAGTAATACCGTTTTATTTCCAGCGCAGTCAGGC 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    684 TTAGTCAATAAATATAATCAAATAAACAGTAATACCGTTTTATTTCCAGCGCAGTCAGGC 743

Qy    661 AGCGGCGTTAAAGTAGCCACTGAAGCGGAAGCGAGACAGTGGCTCAGTGAATTGAATTTA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    744 AGCGGCGTTAAAGTAGCCACTGAAGCGGAAGCGAGACAGTGGCTCAGTGAATTGAATTTA 803

Qy    721 CCGAATAGCTGCCTGAAATCTTATGGATCCGGTTATGTCGTCACCGTTGATCTGACGCCA 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    804 CCGAATAGCTGCCTGAAATCTTATGGATCCGGTTATGTCGTCACCGTTGATCTGACGCCA 863

Qy    781 TTACAAAAAAATGGTTCAGGATATTGATGGTTTAGGCGCGCCGGGAAAAGACTCAAAACTC 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    864 TTACAAAAAAATGGTTCAGGATATTGATGGTTTAGGCGCGCCGGGAAAAGACTCAAAACTC 923

Qy    841 GAAATGGATAACGCCAAATATCAAGCCTGGCAGTCGGGTTTTAAAGCGCAGGAAGAAAAT 900
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    924 GAAATGGATAACGCCAAATATCAAGCCTGGCAGGCGGGTTTTAAAGCGCAGGAAGAAAAT 983

Qy    901 ATGAAAACCACATTACAGACGCTGACGCAAAAATATAGCAATGCCAATTCATTGTACGAC 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    984 ATGAAAACCACATTACAGACGCTGACGCAAAAATATAGCAATGCCAATTCATTGTACGAC 1043

Qy    961 AACCTGGTAAAAGTGCTGAGCAGTACGATAAGTAGCAGCCTGGAAACCGCCAAAAGCTTC 1020
          |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db   1044 AACCTGGTAAAAGTGCTGAGCAGTACGATAAGTAGCAG-CTGGAAACCGCCAAAAGCTTC 1102

Qy   1021 CTGCAAGGATAA 1032
          ||||||||||||
Db   1103 CTGCAAGGATAA 1114
```

```
RESULT   3
STSIPGNS
LOCUS       STSIPGNS     5393 bp    DNA             BCT       04-OCT-1995
DEFINITION  S.typhi spaT, sipB, sipC, sipD and sipA genes.
ACCESSION   X82670
NID         g1009001
VERSION     X82670.1  GI:1009001
KEYWORDS    sipA gene; sipB gene; sipC gene; sipD gene; spaT gene.
SOURCE      Salmonella typhi.
  ORGANISM  Salmonella typhi
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Salmonella.
REFERENCE   1  (bases 1 to 5393)
  AUTHORS   Hermant,D., Menard,R., Arricau,N., Parsot,C. and Popoff,M.Y.
  TITLE     Functional conservation of the Salmonella and Shigella effectors of
            entry into epithelial cells
  JOURNAL   Mol. Microbiol. 17 (4), 781-789 (1995)
  MEDLINE   96111497
REFERENCE   2  (bases 1 to 5393)
  AUTHORS   Popoff,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-1994) M. Popoff, Institut Pasteur, Unite des
            Enterobacteries, 28 rue du Docteur Roux, 75724 Paris Cedex 15,
            FRANCE
FEATURES             Location/Qualifiers
     source          1. .5393
                     /organism="Salmonella typhi"
                     /strain="Ty2"
                     /db_xref="taxon:601"
     gene            43. .540
                     /gene="spaT"
     CDS             43. .540
                     /gene="spaT"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA57987.1"
                     /db_xref="PID:g1009002"
                     /db_xref="GI:1009002"
                     /db_xref="SPTREMBL:Q57260"
                     /translation="MDYQNNVSEERVAEMIWDAVSEGATLKDVHGIPQDMMDGLYAHA
                     YEFYNQGRLDEAETFFRFLCIYDFYNPDYTMGLAAVCQLKKQFQKACDLYAVAFTLLK
                     NDYRPVFFTGQCQLLMRKAAKARQCFELVNERTEDESLRAKALVYLEALKTAETEQHS
                     EQEKE"
     gene            543. .2324
                     /gene="sipB"
     CDS             543. .2324
                     /gene="sipB"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA57988.1"
                     /db_xref="PID:g1009003"
                     /db_xref="GI:1009003"
                     /db_xref="SPTREMBL:Q56134"
                     /translation="MVNDASSISRSGYTQNPRLAEAAFEGVRKNTDFLKAADKAFKDV
                     VATKAGDLKAGTKSGESAINTVGLKPPTDAAREKLSSEGQLTLLLGKLMTLLGDVSLS
                     QLESRLAVWQAMIESQKEMGIQVSKEFQTALGEAQEATDLYEASIKKTDTAKSVYDAA
                     AKKLTQAQNKLQSLDPADPGYAQAEAAVEQAGKEATEAKEALDKATDATVKAGTDAKA
                     KAEKADNILTKFQGTANAASQNQVSQGEQDNLSNVARLTMLMAMFIEIVGKNTEESLQ
                     NDLALFNALQEGRQAEMEKKSAEFQEETRKAEETNRIMGCIGKVLGALLTIVSVVAAV
                     FTGGASLALAAVGLAVMVADEIVKAATGVSFIQQALNPIMEHVLKPLMELIGKAITKA
                     LEGLGVDKKTAEMAGSIVGAIVAAIAMVAVIVVVAVVGKGAAAKLGNALSKMMGETIK
                     KLVPNVLKQLAQNGSKLFTQGMQRITSGLGNVGSKMGLQTNALSKELVGNTLNKVALG
                     MEVTNTAAQSAGGVAEGVFIKNASEALADFMLARFAMDQIQQWLKQSVEIFGENQKVT
                     AELQKAMSSAVQQNADASRFILRQSRA"
     gene            2352. .3581
                     /gene="sipC"
     CDS             2352. .3581
                     /gene="sipC"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA57989.1"
                     /db_xref="PID:g1009004"
                     /db_xref="GI:1009004"
                     /db_xref="SPTREMBL:Q56135"
                     /translation="MLISNVGINPAAYLNNHSVENSSQTASQSVSAKDILNSIGISSS
                     KVSDLGLSPTLSAPAPGVLTQTPGTITSFLKASIQNTDMNQDLNALANNVTTKANEVV
                     QTQLREQQAEVGKFFDISGMSSSAVALLAAANTLMLTLNQADSKLSGKLSLVSFDAAK
                     TTASSMMREGMNALSGSISQSALQLGITGVGAKLEYKGLQNERGALKHNAAKIDKLTT
                     ESHSIKNVLNGQNSVKLGAEGVDSLKSLNMKKTGTDATKNLNDATLKSNAGTSATESL
                     GIKNSNKQISPEHQAILSKRLESVESDIRLEQNTMDMTRIDARKMQMTGDLIMKNSVT
                     VGGIAGASRQYAATQERSEQQISQVNNRVASTASDEARESSRKSTSLIQEMLKIMESI
                     NQSKASALAAIAGNIRA"
     gene            3651. .4670
                     /gene="sipD"
     CDS             3651. .4670
                     /gene="sipD"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA57990.1"
                     /db_xref="PID:g1009005"
                     /db_xref="GI:1009005"
                     /db_xref="SPTREMBL:Q56136"
                     /translation="MLNIQNYSASPHPGIVAERPQTPSASEHAEIAVVPSTTEHRGTD
                     IISLSQAATKIQQAQQTLQSTPPISEENNDERTLARQQLTSSLNALAKSGVSLSAEQN
                     ENLRSTFSARRRPYLALRLWPARTTISDAEIWDMVSQNISAIGDSYLGVYENVVAVYT
                     DFYQAFSDILSKMGGWLSPGKDGNTIKLNVDSLKSEISSLINKYTQINKNTILFPSQT
                     GSGMTTATKAEAEQWIKELNLPDSCLKASGSGYVVLVDTGPLSKMVSDLNGIGSGSAL
                     ELDNAKYQAWQSGFKAQEENLKTTLQTLTQKYSNANSLYDNLVKVLSSTISSSLETAK
                     SFLQG"
     gene            4689. .5393
                     /gene="sipA"
     CDS             4689. .>5393
                     /gene="sipA"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA57991.1"
                     /db_xref="PID:g1009006"
                     /db_xref="GI:1009006"
                     /db_xref="SPTREMBL:Q56137"
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3038 GAAATGGATAACGCCAAATATCAAGCCTGGCAGTCGGGTTTTAAAGCGCAGGAAGAAAAT 3097

Qy     901 ATGAAAACCACATTACAGACGCTGACGCAAAAATATAGCAATGCCAATTCATTGTACGAC 960
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3098 ATGAAAACCACATTACAGACGCTGACGCAAAAATATAGCAATGCCAATTCATTGTACGAC 3157

Qy     961 AACCTGGTAAAAGTGCTGAGCAGTACGATAAGTAGCAGCCTGGAAACCGCCAAAAGCTTC 1020
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3158 AACCTGGTAAAAGTGCTGAGCAGTACGATAAGTAGCAGCCTGGAAACCGCCAAAAGCTTC 3217

Qy    1021 CTGCAAGGATAA 1032
                ||||||||||||
Db    3218 CTGCAAGGATAA 3229


RESULT   2
STU40013
LOCUS       STU40013     3500 bp    DNA             BCT       02-FEB-1996
DEFINITION  Salmonella typhimurium sipD, sipA, and acyl carrier protein (iacP)
            genes, complete cds.
ACCESSION   U40013
NID         g1172126
VERSION     U40013.1  GI:1172126
KEYWORDS    .
SOURCE      Salmonella typhimurium strain=SL1344.
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Salmonella.
REFERENCE   1  (bases 1 to 3500)
  AUTHORS   Kaniga,K., Trollinger,D. and Galan,J.E.
  TITLE     Identification of two targets of the type III protein secretion
            system encoded by the inv and spa loci of Salmonella typhimurium
            that have homology to the Shigella IpaD and IpaA proteins
  JOURNAL   J. Bacteriol. 177 (24), 7078-7085 (1995)
  MEDLINE   96099289
REFERENCE   2  (bases 1 to 3500)
  AUTHORS   Galan,J.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-NOV-1995) Jorge E. Galan, Molecular Gentics and
            Microbiology, School of Medicine, SUNY Stony Brook, Stony Brook, NY
            11794-5222, USA
FEATURES             Location/Qualifiers
     source          1. .3500
                     /organism="Salmonella typhimurium"
                     /strain="SL1344"
                     /db_xref="taxon:602"
     gene            84. .1130
                     /gene="sipD"
     CDS             84. .1130
                     /gene="sipD"
                     /note="homolog of Shigella spp. IpaD"
                     /codon_start=1
                     /transl_table=11
                     /product="SipD"
                     /protein_id="AAA86617.1"
                     /db_xref="PID:g1172127"
                     /db_xref="GI:1172127"
                     /translation="MLNIQNYSASPHPGIVAERPQTPSASEHVETAVVPSTTEHRGTD
                     IISLSQAATKIHQAQQTLQSTPPISEENNDERTLARQQLTSSLNALAKSGVSLSAEQN
                     ENLRSAFSAPTSALFSASPMAQPRTTISDAEIWDMVSQNISAIGDSYLGVYENVVAVY
                     TDFYQAFSDILSKMGGWLLPGKDGNTVKLDVTSLKNDLNSLVNKYNQINSNTVLFPAQ
                     SGSGVKVATEAEARQWLSELNLPNSCLKSYGSGYVVTVDLTPLQKMVQDIDGLGAPGK
                     DSKLEMDNAKYQAWQAGFKAQEENMKTTLQTLTQKYSNANSLYDNLVKVLSSTISSSW
                     KPPKASCKDNRRGY"
     gene            1133. .3187
                     /gene="sipA"
     CDS             1133. .3187
                     /gene="sipA"
                     /note="homolog of Shigella spp. IpaA"
                     /codon_start=1
                     /transl_table=11
                     /product="SipA"
                     /protein_id="AAA86618.1"
                     /db_xref="PID:g1172128"
                     /db_xref="GI:1172128"
                     /translation="MVTSVRTQPPVIMPGMQTEIKTQATNLAANLSAVRAKCHSDAVR
                     EIKGPQLEDFPALIKQASLDALFKCGKDAEALKEVFTNSNNVAGKKAIMEFAGLFRSA
                     LNATSDSPEAKTLLMKVGAEYAAQIIKDGLKEKSAFGPWLPETKKAEAKLENLEKQLL
                     DIIKNNTGGELSKLSTNLVMQEVMPYIASCIEHNFGCTLDPLTRSNLTHLVDKAAAKA
                     VEALDMCPQKLTQEQGTSVGREARHLEMQTLIPLLLRNVFAQIPADKLPDPKIPEPAA
                     GPVPDGGKKAEPTGINININIDSSNHSVDNSKHINNAEPVDNGQRHIDNSNHDNSRKT
                     IDNSRTFIDNSQRNGESHHSTNSSNVSHSHSRVDSTTHQTETAHSASTGAIDHGIAGK
                     IDVTAHATAEAVTNASSESKDGKVVTSEKGTTGETTSFDEVDGVTSKSIIGKPVQATV
                     HGVDDNKQQSQTAEIVNVKPLASQLAGVENVKTDTLQSDTTVITGNKAGTTDNDNSQT
                     DKTGPFSGLKFKQNSFLSTVPSVTNMHSMHFDARETFLGVIRKALEPDTSTPFPVRRA
                     FDGLRAEILPNDTIKSAALKAQCSDIDKHPELKAKMETLKEVITHHPQKEKLAEIALQ
                     FAREAGLTRLKGETDYVLSNVLDGLIGDGSWRAGPAYESYLNKPGVDRVITTVDGLHM
                     QR"
     gene            3206. .3454
                     /gene="iacP"
     CDS             3206. .3454
                     /gene="iacP"
                     /note="IacP"
                     /codon_start=1
                     /transl_table=11
                     /product="acyl carrier protein"
                     /protein_id="AAA86619.1"
                     /db_xref="PID:g1172129"
                     /db_xref="GI:1172129"
                     /translation="MNMDIEARVKKVITSCIAVDVDSINGQTHLVEDLYADSLDLIDI
                     VFGLSEEFDISCNENDLPDMMTFADICRVVKKSLESRV"
BASE COUNT     1048 a    759 c    879 g    814 t
ORIGIN


  Query Match             98.7%;  Score 1018.4;  DB 1;  Length 3500;
  Best Local Similarity   99.8%;  Pred. No. 2.4e-264;
  Matches 1030;  Conservative    0;  Mismatches    1;  Indels    1;  Gaps    1;

Qy       1 ATGCTTAATATTCAAAATTATTCCGCTTCTCCTCATCCGGGGATCGTTGCCGAACGGCCG 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      84 ATGCTTAATATTCAAAATTATTCCGCTTCTCCTCATCCGGGGATCGTTGCCGAACGGCCG 143

Qy      61 CAGACTCCCTCGGCGAGCGAGCACGTCGAGACTGCCGTGGTACCGTCTACCACAGAACAT 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     144 CAGACTCCCTCGGCGAGCGAGCACGTCGAGACTGCCGTGGTACCGTCTACCACAGAACAT 203

Qy     121 CGCGGTACAGATATCATTTCATTATCGCAGGCGGCTACTAAAATCCACCAGGCACAGCAG 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     204 CGCGGTACAGATATCATTTCATTATCGCAGGCGGCTACTAAAATCCACCAGGCACAGCAG 263

Qy     181 ACGCTGCAGTCAACGCCACCGATCTCTGAAGAGAATAATGACGAGCGCACGCTGGCGCGC 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     264 ACGCTGCAGTCAACGCCACCGATCTCTGAAGAGAATAATGACGAGCGCACGCTGGCGCGC 323

Qy     241 CAGCAGTTGACCAGCAGCCTGAATGCGCTGGCGAAGTCCGGCGTGTCATTATCCGCAGAA 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     324 CAGCAGTTGACCAGCAGCCTGAATGCGCTGGCGAAGTCCGGCGTGTCATTATCCGCAGAA 383

Qy     301 CAAAATGAGAACCTGCGGAGCGCGTTTTCTGCGCCGACGTCGGCCTTATTTAGCGCTTCG 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     384 CAAAATGAGAACCTGCGGAGCGCGTTTTCTGCGCCGACGTCGGCCTTATTTAGCGCTTCG 443

Qy     361 CCTATGGCGCAGCCGAGAACAACCATTTCTGATGCTGAGATTTGGGATATGGTTTCCCAA 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     444 CCTATGGCGCAGCCGAGAACAACCATTTCTGATGCTGAGATTTGGGATATGGTTTCCCAA 503

Qy     421 AATATATCGGCGATAGGTGACAGCTATCTGGGCGTTTATGAAAACGTTGTCGCAGTCTAT 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     504 AATATATCGGCGATAGGTGACAGCTATCTGGGCGTTTATGAAAACGTTGTCGCAGTCTAT 563

Qy     481 ACCGATTTTTATCAGGCCTTCAGTGATATTCTTTCCAAAATGGGAGGCTGGTTATTACCA 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     564 ACCGATTTTTATCAGGCCTTCAGTGATATTCTTTCCAAAATGGGAGGCTGGTTATTACCA 623
```

```
          Massachusetts General Hospital, Harvard Medical School, Fruit
          Street, Boston, MA 02114, USA
FEATURES             Location/Qualifiers
     source          1. .3441
                     /organism="Salmonella typhimurium"
                     /strain="SL1344"
                     /db_xref="taxon:602"
                     /map="59 min"
     gene            1. .870
                     /gene="sspB"
     CDS             <1. .870
                     /gene="sspB"
                     /note="secreted protein"
                     /codon_start=1
                     /transl_table=11
                     /function="invasion determinant"
                     /product="SspB"
                     /protein_id="AAC43546.1"
                     /db_xref="PID:g1117812"
                     /db_xref="GI:1117812"
                     /translation="RKAEETNRIMGCIGKVLGALLTIVSVVAAVFTGGASLALAAVGL
                     AVMVADEIVKAATGVSFIQQALNPIMEHVLKPLMELIGKAITKALEGLGVDKKTAEMA
                     GSIVGAIVAAIAMVAVIVVVAVVGKGAAAKLGNALSKMMGETIKKLVPNVLKQLAQNG
                     SKLFTQGMQRITSGLGNVGSKMGLQTNALSKELVGNTLNKVALGMEVTNTAAQSAGGV
                     AEGVFIKNASEALADFMLARFAMDQIQQWLKQSVEIFGENQKVTAELQKAMSSAVQQN
                     ADASRFILRQSRA"
     gene            898. .2127
                     /gene="sspC"
     CDS             898. .2127
                     /gene="sspC"
                     /note="secreted protein"
                     /codon_start=1
                     /transl_table=11
                     /function="invasion determinant"
                     /evidence=experimental
                     /product="SspC"
                     /protein_id="AAC43547.1"
                     /db_xref="PID:g1117813"
                     /db_xref="GI:1117813"
                     /translation="MLISNVGINPAAYLNNHSVENSSQTASQSVSAKDILNSIGISSS
                     KVSDLGLSPTLSAPAPGVLTQTPGTITSFLKASIQNTDMNQDLNALANNVTTKANEVV
                     QTQLREQQAEVGKFFDISGMSSSAVALLAAANTLMLTLNQADSKLSGKLSLVSFDAAK
                     TTASSMMREGMNALSGSISQSALQLGITGVGAKLEYKGLQNERGALKHNAAKIDKLTT
                     ESHSIKNVLNGQNSVKLGAEGVDSLKSLNMKKTGTDATKNLNDATLKSNAGTSATESL
                     GIKDSNKQISPEHQAILSKRLESVESDIRLEQNTMDMTRIDARKMQMTGDLIMKNSVT
                     VGGIAGASGQYAATQERSEQQISQVNNRVASTASDEARESSRKSTSLIQEMLKTMESI
                     NQSKASALAAIAGNIRA"
     gene            2198. .3229
                     /gene="sspD"
     CDS             2198. .3229
                     /gene="sspD"
                     /note="secreted protein"
                     /codon_start=1
                     /transl_table=11
                     /function="invasion determinant"
                     /evidence=experimental
                     /product="SspD"
                     /protein_id="AAC43548.1"
                     /db_xref="PID:g1117814"
                     /db_xref="GI:1117814"
                     /translation="MLNIQNYSASPHPGIVAERPQTPSASEHVETAVVPSTTEHRGTD
                     IISLSQAATKIHQAQQTLQSTPPISEENNDERTLARQQLTSSLNALAKSGVSLSAEQN
                     ENLRSAFSAPTSALFSASPMAQPRTTISDAEIWDMVSQNISAIGDSYLGVYENVVAVY
                     TDFYQAFSDILSKMGGWLLPGKDGNTVKLDVTSLKNDLNSLVNKYNQINSNTVLFPAQ
                     SGSGVKVATEAEARQWLSELNLPNSCLKSYGSGYVVTVDLTPLQKMVQDIDGLGAPGK
                     DSKLEMDNAKYQAWQSGFKAQEENMKTTLQTLTQKYSNANSLYDNLVKVLSSTISSSL
                     ETAKSFLQG"
     gene            3248. .3441
                     /gene="sspA"
     CDS             3248. .>3441
                     /gene="sspA"
                     /note="secreted protein"
                     /codon_start=1
                     /transl_table=11
                     /function="unknown"
                     /product="SspA"
                     /protein_id="AAC43549.1"
                     /db_xref="PID:g1117815"
                     /db_xref="GI:1117815"
                     /translation="MVTSVRTQPPVIMPGMQTEIKTQATNLAANLSAVRESATATLSG
                     EIKGPQLEDFPALIKQASLDA"
BASE COUNT      978 a    759 c    925 g    779 t
ORIGIN
```

Query Match 100.0%; Score 1032; DB 2; Length 3441;
Best Local Similarity 100.0%; Pred. No. 5.1e-268;
Matches 1032; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 ATGCTTAATATTCAAAATTATTCCGCTTCTCCTCATCCGGGGATCGTTGCCGAACGGCCG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2198 ATGCTTAATATTCAAAATTATTCCGCTTCTCCTCATCCGGGGATCGTTGCCGAACGGCCG 2257

Qy      61 CAGACTCCCTCGGCGAGCGAGCACGTCGAGACTGCCGTGGTACCGTCTACCACAGAACAT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2258 CAGACTCCCTCGGCGAGCGAGCACGTCGAGACTGCCGTGGTACCGTCTACCACAGAACAT 2317

Qy     121 CGCGGTACAGATATCATTTCATTATCGCAGGCGGCTACTAAAATCCACCAGGCACAGCAG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2318 CGCGGTACAGATATCATTTCATTATCGCAGGCGGCTACTAAAATCCACCAGGCACAGCAG 2377

Qy     181 ACGCTGCAGTCAACGCCACCGATCTCTGAAGAGAATAATGACGAGCGCACGCTGGCGCGC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2378 ACGCTGCAGTCAACGCCACCGATCTCTGAAGAGAATAATGACGAGCGCACGCTGGCGCGC 2437

Qy     241 CAGCAGTTGACCAGCAGCCTGAATGCGCTGGCGAAGTCCGGCGTGTCATTATCCGCAGAA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2438 CAGCAGTTGACCAGCAGCCTGAATGCGCTGGCGAAGTCCGGCGTGTCATTATCCGCAGAA 2497

Qy     301 CAAAATGAGAACCTGCGGAGCGCGTTTTCTGCGCCGACGTCGGCCTTATTTAGCGCTTCG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2498 CAAAATGAGAACCTGCGGAGCGCGTTTTCTGCGCCGACGTCGGCCTTATTTAGCGCTTCG 2557

Qy     361 CCTATGGCGCAGCCGAGAACAACCATTTCTGATGCTGAGATTTGGGATATGGTTTCCCAA 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2558 CCTATGGCGCAGCCGAGAACAACCATTTCTGATGCTGAGATTTGGGATATGGTTTCCCAA 2617

Qy     421 AATATATCGGCGATAGGTGACAGCTATCTGGGCGTTTATGAAAACGTTGTCGCAGTCTAT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2618 AATATATCGGCGATAGGTGACAGCTATCTGGGCGTTTATGAAAACGTTGTCGCAGTCTAT 2677

Qy     481 ACCGATTTTTATCAGGCCTTCAGTGATATTCTTTCCAAAATGGGAGGCTGGTTATTACCA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2678 ACCGATTTTTATCAGGCCTTCAGTGATATTCTTTCCAAAATGGGAGGCTGGTTATTACCA 2737

Qy     541 GGTAAGGACGGTAATACCGTTAAGCTAGATGTTACCTCACTCAAAAATGATTTAAACAGT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2738 GGTAAGGACGGTAATACCGTTAAGCTAGATGTTACCTCACTCAAAAATGATTTAAACAGT 2797

Qy     601 TTAGTCAATAAATATAATCAAATAAACAGTAATACCGTTTTATTTCCAGCGCAGTCAGGC 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2798 TTAGTCAATAAATATAATCAAATAAACAGTAATACCGTTTTATTTCCAGCGCAGTCAGGC 2857

Qy     661 AGCGGCGTTAAAGTAGCCACTGAAGCGGAAGCGAGACAGTGGCTCAGTGAATTGAATTTA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2858 AGCGGCGTTAAAGTAGCCACTGAAGCGGAAGCGAGACAGTGGCTCAGTGAATTGAATTTA 2917

Qy     721 CCGAATAGCTGCCTGAAATCTTATGGATCCGGTTATGTCGTCACCGTTGATCTGACGCCA 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2918 CCGAATAGCTGCCTGAAATCTTATGGATCCGGTTATGTCGTCACCGTTGATCTGACGCCA 2977

Qy     781 TTACAAAAAAATGGTTCAGGATATTGATGGTTTAGGCGCGCCGGGAAAAGACTCAAAACTC 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2978 TTACAAAAAAATGGTTCAGGATATTGATGGTTTAGGCGCGCCGGGAAAAGACTCAAAACTC 3037

Qy     841 GAAATGGATAACGCCAAATATCAAGCCTGGCAGTCGGGTTTTAAAGCGCAGGAAGAAAAT 900
```

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 1999, 17:58:36 ; Search time 2414.69 Seconds
(without alignments)
1359.214 Million cell updates/sec

Title: US-09-068-804-3
Perfect score: 1032
Sequence: 1 ATGCTTAATATTCAAAATTA.....AAAGCTTCCTGCAAGGATAA 1032

Scoring table: IDENTITY_NUC

Searched: 679419 seqs, 1590154680 residues

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_st:*
14: gb_sts:*
15: gb_sy:*
16: gb_un:*
17: gb_vi:*
18: em_fun:*
19: em_htg:*
20: em_hum1:*
21: em_hum2:*
22: em_in:*
23: em_om:*
24: em_or:*
25: em_ov:*
26: em_pat:*
27: em_ph:*
28: em_pl:*
29: em_ro:*
30: em_sts:*
31: em_sy:*
32: em_un:*
33: em_vi:*
34: gb_htg1:*
35: gb_htg2:*
36: gb_in1:*
37: gb_in2:*
38: em_ba1:*
39: em_ba2:*
40: em_hum3:*
41: em_hum4:*
42: gb_pr4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1032	100.0	3441	2	STU30491	U30491 Salmonella
2	1018.4	98.7	3500	1	STU40013	U40013 Salmonella
3	818	79.3	5393	1	STSIPGNS	X82670 S.typhi spa
4	101.6	9.8	6869	1	SDIPABCD	X60777 Shigella dy
5	101.6	9.8	7663	1	SFPVR2	X15319 Shigella fl
6	101.6	9.8	5255	1	SHFIPABCD	J04117 S.flexneri
7	64.6	6.3	4761	1	W10PPIBCA	M34849 Plasmid pWR
c 8	48.8	4.7	7218	5	I66494	I66494 Sequence 14
9	39.2	3.8	95765	8	ATAC006528	AC006528 Arabidops
c 10	39.2	3.8	152637	34	AC006734	AC006734 Caenorhab
c 11	39.2	3.8	38455	36	CELT03G11	U41272 Caenorhabdi
c 12	38.4	3.7	4395	36	PFARHOPH3	M65059 Plasmodium
c 13	37.6	3.6	110000	34	CEY102G3_2	Continuation (3 of
c 14	37.6	3.6	110000	34	CEY113G7_0	AL031113 Caenorhab
c 15	37.6	3.6	192849	34	CEY60A3	AL021574 Caenorhab
c 16	37.4	3.6	202734	34	PFMAL13P1	AL049180 Plasmodiu
17	37	3.6	236120	17	AF063866	AF063866 Melanoplu
c 18	36.8	3.6	110000	34	CEY102F5_0	AL022276 Caenorhab
c 19	36.8	3.6	102195	34	CEY71H9	AL021575 Caenorhab
20	36.8	3.6	116525	34	HS688G8	AL031671 Homo sapi
c 21	36.8	3.6	37872	36	CET04F8	Z66565 Caenorhabdi
c 22	36.6	3.5	118758	9	HS450C20	Z84720 Human DNA s
23	36.4	3.5	204104	34	CNS0000A	AL049828 Homo sapi
24	36.4	3.5	300172	35	AC005308	AC005308 Plasmodiu
25	36.4	3.5	36822	36	CELC25E10	U50311 Caenorhabdi
26	36	3.5	1809	1	BSMBSUFI	X51515 Bacillus su
c 27	35.4	3.4	181571	42	AC007103	AC007103 Homo sapi
28	35.2	3.4	115359	8	ATAC003028	AC003028 Arabidops
29	35.2	3.4	505	14	GGU97544	U97544 Gallus gall
c 30	35.2	3.4	114596	34	HS1063B2	AL035683 Homo sapi
c 31	35.2	3.4	181634	35	AC007745	AC007745 Homo sapi
32	35	3.4	85855	8	AC002311	AC002311 Arabidops
c 33	35	3.4	99053	8	AC005292	AC005292 Genomic s
34	35	3.4	4958	10	HSPAPPA	X68280 H.sapiens P
35	35	3.4	8400	11	HSU28727	U28727 Human pregn
36	35	3.4	110000	34	PFMAL4P1_0	AL034557 Plasmodiu
c 37	34.8	3.4	6478	7	AB023040	AB023040 Arabidops
38	34.8	3.4	129756	10	HS875H10	AL034350 Human DNA
c 39	34.8	3.4	230278	17	MCU68299	U68299 Mouse cytom
40	34.8	3.4	163142	42	AC005877	AC005877 Homo sapi
41	34.6	3.4	183871	35	AC007365	AC007365 Homo sapi
c 42	34.6	3.4	200218	35	AC007690	AC007690 Homo sapi
c 43	34.6	3.4	9496	36	CELF56D3	U28942 Caenorhabdi
c 44	34.4	3.3	13341	8	OLRKLP1	U72726 Oryza longi
c 45	34.4	3.3	96608	9	HS308I13	Z99291 Human DNA s

ALIGNMENTS

RESULT 1
STU30491
LOCUS STU30491 3441 bp DNA BCT 19-JAN-1996
DEFINITION Salmonella typhimurium SspA (sspA) and SspB (sspB) genes, partial cds, and SspC (sspC) and SspD (sspD) genes, complete cds.
ACCESSION U30491
NID g1117811
VERSION U30491.1 GI:1117811
KEYWORDS .
SOURCE Salmonella typhimurium.
ORGANISM Salmonella typhimurium
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Salmonella.
REFERENCE 1 (bases 1 to 3441)
AUTHORS Hueck,C.J., Hantman,M.J., Bajaj,V., Johnston,C., Lee,C.A. and Miller,S.I.
TITLE Salmonella typhimurium secreted invasion determinants are homologous to Shigella Ipa proteins
JOURNAL Mol. Microbiol. 18 (3), 479-490 (1995)
MEDLINE 96342378
REFERENCE 2 (bases 1 to 3441)
AUTHORS Miller,S.I.
TITLE Direct Submission
JOURNAL Submitted (28-JUN-1995) Samuel I. Miller, Infectious Disease Unit,

This Page Blank (uspto)

KW sublocalisation; gene tagging; tissue typing.
OS Synthetic.
PN WO9300353-A.
PD 07-JAN-1993.
PF 19-JUN-1992; U05222.
PR 20-JUN-1991; US-716831.
PR 12-FEB-1992; US-837195.
PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
PI Adams MD, Venter JC;
DR WPI; 93-036325/04.
PT Particular expressed sequence tags from human CDNA - corresponds
PT to transcription prods. of genes, useful for tagging genes,
PT mapping chromosomes and tissue typing
PS Claim 3; Page 128; 199pp; English.
CC This sequence represents an EST (expressed sequence tag) ESTs are markers
CC for human genes actually transcribed in vivo. Unlike the random genomic
CC DNA sequence tagged sites (STSs), ESTs point directly to expressed genes.
CC The use of ESTs could facilitate the tagging of most expressed human
CC genes within a few years at a fraction of the cost of complete genomic
CC sequencing. Using PCR primers Q39419-Q39580 (sequences designed
CC from the ESTs) sublocalisation of an EST can be achieved with panels of
CC fragments from specific chromosomes or pools of large genomic clones in
CC an analogous manner. This sequence represents EST00111.
SQ Sequence 529 BP; 141 A; 116 C; 99 G; 173 T;

Query Match 3.1%; Score 32.2; DB 1; Length 529;
Best Local Similarity 52.6%; Pred. No. 1.5;
Matches 70; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

```
Qy     824 GAAAAGACTCAAAACTCGAAATGGATAACGCCAAATATCAAGCCTGGCAGTCGGGTTTTA 883
           | ||| |  | |||||   | |        ||| || ||   | |  || | |  ||
Db     419 GCAAACAAACCAAACTGCCAGTCTCATTTCCCACATTTCTTACTTAAGAGAAGAGAAGTA 360

Qy     884 AAGCGCAGGAAGAAAATATGAAAACCACATTACAGACGCTGACGCAAAAATATAGCAATG 943
           ||    |||||||| | ||  |      |||| ||| |  |    ||| |||  ||   |
Db     359 AATGAAAGGAAGAAGAAATAGATTTGTAATTAAAGATGTGGCAAAAAAGATAGGGCTGAG 300

Qy     944 CCAATTCATTGTA 956
           ||| |||| | ||
Db     299 CCAGTTCAATTTA 287
```

RESULT 15
Q59170/c
ID Q59170 standard; cDNA; 529 BP.
AC Q59170;
DI 16-MAR-1994 (first entry)
DE Human brain Expressed Sequence Tag EST00111.
KW Gene transcription product; genetic markers; tagging; in vivo;
KW transcription; mapping; locations; chromosomes; chromosomal; ss.
OS Homo sapiens.
PN WO9316178-A.
PD 19-AUG-1993.
PF 12-FEB-1993; U01294.
PR 12-FEB-1992; US-837195.
PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
PI Adams MD, Moreno RF, Venter CJ;
DR WPI; 93-272882/34.
PT Enriched oligonucleotides and corresp. sequences - used as
PT markers for human genes transcribed in-vivo, facilitate tagging
PI of most human genes
PS Example 1; Page 132; 500pp; English.
CC The Expressed Sequence Tag was isolated from a human brain cDNA
CC library as part of a large set of ESTs which can be used as markers
CC for human genes transcribed in vivo. They can be used to facilitate
CC tagging of most human genes, for mapping locations of expressed genes
CC on chromosomes, for individual or forensic identification, for mapping
CC locations of disease-associated genes, for identification of tissue
CC type, and for prepn. of antisense sequences, probes and constructs.
CC EST00111 has a "poor" coding probability as evaluated using the
CC coding-region prediction program CRM and hybridises to chromosome 11.
CC See also Q59041-Q61440.

SQ Sequence 529 BP; 141 A; 116 C; 99 G; 173 T;

Query Match 3.1%; Score 32.2; DB 1; Length 529;
Best Local Similarity 52.6%; Pred. No. 1.5;
Matches 70; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

```
Qy     824 GAAAAGACTCAAAACTCGAAATGGATAACGCCAAATATCAAGCCTGGCAGTCGGGTTTTA 883
           | ||| |  | |||||   | |        ||| || ||   | |  || | |  ||
Db     419 GCAAACAAACCAAACTGCCAGTCTCATTTCCCACATTTCTTACTTAAGAGAAGAGAAGTA 360

Qy     884 AAGCGCAGGAAGAAAATATGAAAACCACATTACAGACGCTGACGCAAAAATATAGCAATG 943
           ||    |||||||| | ||  |      |||| ||| |  |    ||| |||  ||   |
Db     359 AATGAAAGGAAGAAGAAATAGATTTGTAATTAAAGATGTGGCAAAAAAGATAGGGCTGAG 300

Qy     944 CCAATTCATTGTA 956
           ||| |||| | ||
Db     299 CCAGTTCAATTTA 287
```

Search completed: October 5, 1999, 17:59:04
Job time: 11386 sec

KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KW computer readable medium; vaccine; pharmaceutical composition; ds.
OS Streptococcus pneumoniae.
PN WO9818931-A2.
PD 07-MAY-1998.
PF 30-OCT-1997; U19588.
PR 31-OCT-1996; US-029960.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M,
PI Kunsch CA, Rosen CA;
DR WPI; 98-272225/24.
PT Computer-readable medium with recorded Streptococcus pneumoniae
PT polynucleotide sequences - useful in diagnostic kits and assays, and
PT pharmaceutical compositions and vaccines for Streptococcus
PT pneumoniae
PS Claim 1; Page 1209-1211; 1409pp; English.
CC The present invention describes a computer readable medium which has
CC the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded
CC on it, or a representative fragment or a sequence at least 95% identical
CC to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1
CC to 391 (V52134 to V52524) are genomic fragments from Streptococcus
CC pneumoniae. The present invention also describes an isolated nucleic acid
CC molecule encoding a homologue of any of the fragments of the S.pneumoniae
CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced
CC by a process comprising: (a) screening a genomic DNA library using as a
CC probe a target sequence defined by any of the sequences in SEQ ID NO:1
CC to 391, identifying members of the library which contain sequences
CC that hybridise to the target sequence and isolating the nucleic acid
CC molecules from the members; or (b) isolating mRNA, DNA or cDNA produced
CC from an organism, amplifying nucleic acid molecules whose nucleotide
CC sequence is homologous to amplification primers derived from the
CC fragment of the S. pneumoniae genome to prime the amplification and
CC isolating the amplified sequences. The computer readable medium can be
CC used in a computer-based system for identifying fragments of the
CC S. pneumoniae genome of commercial importance, or expression modulating
CC fragments of the S. pneumoniae genome. Products from the present
CC invention can be used in diagnosis kits and assays, and pharmaceutical
CC compositions and vaccines for S. pneumoniae.
SQ Sequence 3902 BP; 1138 A; 714 C; 834 G; 1216 T;

Query Match 3.2%; Score 32.8; DB 1; Length 3902;
Best Local Similarity 48.4%; Pred. No. 2.9;
Matches 91; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

Qy 258 CCTGAATGCGCTGGCGAAGTCCGGCGTGTCATTATCCGCAGAACAAAATGAGAACCTGCG 317
Db 383 CCAGATTCAGCTGAAGATGTCAACCCAATCGTTATGAACGGTATTTCAACAGGAACTTCC 324

Qy 318 GAGCGCGTTTTCTGCGCCGACGTCGGCCTTATTTAGCGCTTCGCCTATGGCGCAGCCGAG 377
Db 323 AACCATGACTTCTTCTCTCAAGTCGGAAATGGTTACCTTCTTGGCGAAGTTGAAGCCATG 264

Qy 378 AACAACCATTTCTGATGCTGAGATTTGGGATATGGTTTCCCAAAATATATCGGCGATAGG 437
Db 263 CAAGATGATGACTACAACTACGGTTTGCACTAATACTCTTCGAAAATCAAATTCAAACCA 204

Qy 438 TGACAGCT 445
Db 203 CGTCAGCT 196

RESULT 13
V74474
ID V74474 standard; DNA; 5132 BP.
AC V74474;
DT 16-MAR-1999 (first entry)
DE Staphylococcus aureus contig SEQ ID #163.
KW Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW skin infection; surgical wound infection; scalded skin syndrome;
KW toxic shock syndrome; ds.
OS Staphylococcus aureus.

FH Key Location/Qualifiers
FT misc_feature 181. .240
FT /*tag= a
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 1981. .2040
FT /*tag= b
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 3781. .3840
FT /*tag= c
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
PN EP-786519-A2.
PD 30-JUL-1997.
PF 07-JAN-1997; 100117.
PR 05-JAN-1996; US-009861.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI Rosen CA;
DR WPI; 97-374922/35.
PT Polynucleotide(s) and proteins derived from Staphylococcus aureus -
PT stored on computer readable medium and used in the production of
PT anti-S.aureus vaccines
PS Claim 1; Page 797-800; 3271pp; English.
CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC of the invention. The DNA sequences are recorded on a computer readable
CC medium, preferably selected from a floppy or hard disk, random access
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC the S.aureus DNA sequences allows putative functions to be assigned so
CC that protein-encoding or regulatory regions of commercial, therapeutic or
CC industrial importance can be obtained. Specifically, sequences which are
CC likely to encode antigens have been identified and these polypeptides can
CC be used in a vaccine composition against S.aureus infection. The
CC polypeptides can also be used in a kit for the immunodetection of
CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC skin and surgical wound infections, scalded skin syndrome, toxic shock
CC syndrome, etc. Organisms transformed with the DNA sequences can be used
CC for recombinant production of the polypeptides. The new DNA sequences
CC (and their fragments) are useful as primers or probes for isolating
CC homologues of any of the S.aureus DNA sequences contained on the
CC computer readable medium.
SQ Sequence 5132 BP; 1589 A; 875 C; 788 G; 1694 T;

Query Match 3.1%; Score 32.4; DB 1; Length 5132;
Best Local Similarity 60.0%; Pred. No. 4.6;
Matches 54; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

Qy 557 CCGTTAAGCTAGATGTTACCTCACTCAAAAATGATTTAAACAGTTTAGTCAATAAATATA 616
Db 3533 CCTTTAAATCATTGCGTGCTTCACTGAAAATTTGTATAAAGATTTAAGTCATTACGTAAC 3592

Qy 617 ATCAAATAAACAGTAATACCGTTTTATTTC 646
Db 3593 ATCACATAAAATACATTTCTATACTATTCC 3622

RESULT 14
Q39758/c
ID Q39758 standard; DNA; 529 BP.
AC Q39758;
DT 20-MAY-1993 (first entry)
DE Expressed Sequence Tag human gene marker EST00111.
KW expressed sequence tag; human genome project; chromosome;
KW human gene sequencing; PCR mapping; somatic cell hybrids;

```
CC   coli hosts was not attained using large expression constructs from the
CC   toxin A gene, usable levels of recombinant protein were obtd. by
CC   isolating induced protein from a series of smaller pMAL expression
CC   constructs that spanned the entire toxin A gene.  Toxin A is used in
CC   antitoxin and vaccine prodn.
SQ   Sequence   8133 BP;    3222 A;    978 C;    1233 G;    2700 T;


  Query Match             3.2%;  Score 33.4;  DB 1;  Length 8133;
  Best Local Similarity   49.2%;  Pred. No. 2.8;
  Matches   88;  Conservative    0;  Mismatches   91;  Indels    0;  Gaps    0;

Qy     824 GAAAAGACTCAAAACTCGAAATGGATAACGCCAAATATCAAGCCTGGCAGTCGGGTTTTA 883
           ||| | |   ||||||  ||   || |   |   | | || |  ||   |   |   | |
Db     177 GAATAAATATAAAACTTCAAGCAGAAATAGAGCACTCTCTAATCTAAAAAAAGATATATT 236

Qy     884 AAGCGCAGGAAGAAAATATGAAAACCACATTACAGACGCTGACGCAAAAATATAGCAATG 943
           ||  | ||  ||        | ||||   || ||||  | |||  | |||||  |  || |
Db     237 AAAAGAAGTAATTCTTATTAAAAATTCCAATACAAGCCCTGTAGAAAAAAATTTACATTT 296

Qy     944 CCAATTCATTGTACGACAACCTGGTAAAAGTGCTGAGCAGTACGATAAGTAGCAGCCTG 1002
             ||  || |   ||  ||    || | | ||||    | |||   ||  |   | |||
Db     297 TGTATGGATAGGTGGAGAAGTCAGTGATATTGCTCTTGAATACATAAAACAATGGGCTG 355


RESULT  10
V30560
ID   V30560 standard; DNA; 8133 BP.
AC   V30560;
DT   07-DEC-1998  (first entry)
DE   Clostridium difficile toxin A gene coding region.
KW   Antitoxin; vaccine; cytotoxin; toxin A; intoxication; immunogen;
KW   pseudomembranous enterocolitis; ds.
OS   Clostridium difficile.
PN   WO9808540-A1.
PD   05-MAR-1998.
PF   28-AUG-1997; U15394.
PR   28-AUG-1996; US-704159.
PA   (OPHI-) OPHIDIAN PHARM INC.
PI   Thalley BS, Williams JA;
DR   WPI; 98-230234/20.
DR   P-PSDB; W68387.
PT   Host cell containing recombinant expression vector encoding
PT   Clostridium botulinum type B or E toxin  - useful to treat humans
PT   and other animals at risk of intoxication with clostridial toxin
PS   Example 15; Page 220-230; 428pp; English.
CC   This is the coding region of the Clostridium difficile toxin A
CC   gene.  It codes for a 2710-amino acid toxin A protein (see W68387).
CC   Toxin A is a potent cytotoxin that plays a direct role in damaging
CC   gastrointestinal tissues.  Severe cases of C. difficile
CC   intoxication result in pseudomembranous colitis.  This would be
CC   prevented by neutralising the effects of toxin A in the
CC   gastrointestinal tract.  Examples are provided of the production
CC   of recombinant C. difficile toxin A in host cells and of the in
CC   vivo neutralisation of toxin A by antibodies against recombinant
CC   toxin A polypeptides.  The invention specifically relates to
CC   recombinant proteins derived from Clostridium botulinum toxins
CC   (see W68389-400) and their use as immunogens for the production of
CC   vaccines and antitoxins.
SQ   Sequence   8133 BP;    3222 A;    978 C;    1233 G;    2700 T;


  Query Match             3.2%;  Score 33.4;  DB 1;  Length 8133;
  Best Local Similarity   49.2%;  Pred. No. 2.8;
  Matches   88;  Conservative    0;  Mismatches   91;  Indels    0;  Gaps    0;

Qy     824 GAAAAGACTCAAAACTCGAAATGGATAACGCCAAATATCAAGCCTGGCAGTCGGGTTTTA 883
           ||| | |   ||||||  ||   || |   |   | | || |  ||   |   |   | |
Db     177 GAATAAATATAAAACTTCAAGCAGAAATAGAGCACTCTCTAATCTAAAAAAAGATATATT 236

Qy     884 AAGCGCAGGAAGAAAATATGAAAACCACATTACAGACGCTGACGCAAAAATATAGCAATG 943
           ||  | ||  ||        | ||||   || ||||  | |||  | |||||  |  || |
Db     237 AAAAGAAGTAATTCTTATTAAAAATTCCAATACAAGCCCTGTAGAAAAAAATTTACATTT 296

Qy     944 CCAATTCATTGTACGACAACCTGGTAAAAGTGCTGAGCAGTACGATAAGTAGCAGCCTG 1002
             ||  || |   ||  ||    || | | ||||    | |||   ||  |   | |||
Db     297 TGTATGGATAGGTGGAGAAGTCAGTGATATTGCTCTTGAATACATAAAACAATGGGCTG 355


RESULT  11
X00225
ID   X00225 standard; DNA; 1560 BP.
AC   X00225;
DT   23-MAR-1999  (first entry)
DE   Clostridium chauvoei blackleg microbe flagellin gene.
KW   Clostridium chauvoei; blackleg microbe; flagellin; vaccine;
KW   diagnosis; infection-preventing antigen; ds.
OS   Clostridium chauvoei.
FH   Key             Location/Qualifiers
FT   RBS             141. .146
FT                   /*tag= a
FT   CDS             156. .1397
FT                   /*tag= b
FT   primer_bind     complement (627. .646)
FT                   /*tag= c
FT                   /note= "PCR primer 1"
FT   primer_bind     1123. .1142
FT                   /*tag= d
FT                   /note= "PCR primer 2"
PN   J11000173-A.
PD   06-JAN-1999.
PF   12-JUN-1997; 169698.
PR   12-JUN-1997; JP-169698.
PA   (NORI-) NORINSUISANSHO DOBUTSU IYAKUHIN KENSA SH.
DR   WPI; 99-124393/11.
DR   P-PSDB; W89678.
PT   New blackleg microbe flagellin gene - useful for vaccine against
PT   blackleg and in method for diagnosing blackleg
PS   Claim 1; Page 6-8; 10pp; Japanese.
CC   The present sequence encodes flagellin from the blackleg microbe
CC   (Clostridium chauvoei). The present invention also describes: (a) a
CC   blackleg vaccine containing the blackleg microbe flagellin gene or an
CC   expression protein of the flagellin gene as the active component; (b) a
CC   method for diagnosing blackleg using PCR comprising primer 1 (X00226)
CC   and primer 2 (X00227), and amplification of the product is analysed by
CC   electrophoresis; and (c) a method for the inspection of the
CC   specification of the blackleg vaccine or the diagnostic solution for
CC   blackleg comprising the flagellin gene or an expression protein of the
CC   flagellin gene. The flagellin gene is capable of expressing a protein
CC   having a function of infection-preventing antigen.
SQ   Sequence   1560 BP;    664 A;    205 C;    285 G;    406 T;


  Query Match             3.2%;  Score 33.2;  DB 1;  Length 1560;
  Best Local Similarity   53.0%;  Pred. No. 1.3;
  Matches   71;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy     508 ATTCTTTCCAAAATGGGAGGCTGGTTATTACCAGGTAAGGACGGTAATACCGTTAAGCTA 567
           ||||||   | |||| |||  |  | | ||| |       ||| ||  | |||     ||
Db     411 ATTCTTCAAAGAATGAGAGAATTATCAGTACAATCAGCTAACGATACAAACGTAGCTGTA 470

Qy     568 GATGTTACCTCACTCAAAAATGATTTAAACAGTTTAGTCAATAAATATAATCAAATAAAC 627
           |||   ||  |  || || ||||  |||||   |||    |  ||   ||   ||||
Db     471 GATAGAACTGCTATCCAAGATGAAATAAACTCATTAACAGAAGAAATAAACAGAATATCA 530

Qy     628 AGTAATACCGTTTT 641
           |  |||| |  ||
Db     531 GGAGATACTGAATT 544


RESULT  12
V52345/c
ID   V52345 standard; DNA; 3902 BP.
AC   V52345;
DT   23-OCT-1998  (first entry)
DE   Streptococcus pneumoniae genome fragment SEQ ID NO:212.
```

```
Db      67 ACTACAAGAGCACGCACACGAGCATCCAGAGCACGACC 104


RESULT   7
V65242/c
ID   V65242 standard; DNA; 3189 BP.
AC   V65242;
DT   24-DEC-1998  (first entry)
DE   DNA encoding three S. pneumoniae proteins with different functions.
KW   Streptococcus pneumoniae protein; recombinant; gene expression;
KW   DNA chip; virulence; antibody; infection; detection; treatment; ss.
OS   Streptococcus pneumoniae.
PN   WO9826072-A1.
PD   18-JUN-1998.
PF   09-DEC-1997; U22578.
PR   13-DEC-1996; US-036281.
PA   (ELIL ) LILLY & CO ELI.
PI   Baltz RH, Burgett SG, Dehoff BS, Hoskins JA, Jaskunas SR,
PI   Mills BJ, Norris FH, Peery RB, Rockey PK, Rosteck PR,
PI   Skatrud PL, Smith MC, Solenberg PJ, Treadway PJ,
PI   Young Bellido ML;
DR   WPI; 98-348529/30.
DR   P-PSDB; W80666, W80667, W80668.
PT   Streptococcus pneumoniae nucleic acid sequences - used in DNA chips
PT   for evaluating gene expression, and identification of virulence
PT   genes
PS   Claim 1; Pages 94-96; 333pp; English.
CC   This DNA encodes S. pneumoniae nitrate reduction, formate dependent
CC   protein, lacRepressor and a nucleotide reductase. The invention provides
CC   DNA sequences (V65201 to V65304) from the S. pneumoniae genome and
CC   corresponding protein sequences (W80605 to W80728). A recombinant host
CC   containing a vector comprising any of the above nucleic acids can be used
CC   for the recombinant expression of the protein sequences. The invention
CC   also provides a DNA chip having arrayed on it at least 15 base pair
CC   fragment of any one or more of these DNA sequences. The DNA chip can be
CC   used methods for evaluating gene expression in S. pneumoniae and for
CC   identifying virulence genes in S. pneumoniae. Antibodies that selectively
CC   bind to the above proteins or peptide fragments can be used to treat
CC   S. pneumoniae infection. The antibodies can also be used to detect
CC   S. pneumoniae cells.
SQ   Sequence   3189 BP;    956 A;    547 C;    615 G;    1071 T;


  Query Match             3.3%;  Score 33.6;  DB 1;  Length 3189;
  Best Local Similarity   51.3%;  Pred. No. 1.5;
  Matches   78;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy    258 CCTGAATGCGCTGGCGAAGTCCGGCGTGTCATTATCCGCAGAACAAAATGAGAACCTGCG 317
          || || |  |||| || |||   |   || ||||  | | |    |  || | || |
Db   1840 CCAGATTCAGCTGAAGATGTCAACCCAATCGTTATGAACGGTATTTCAACAGGAACTTCC 1781

Qy    318 GAGCGCGTTTTCTGCGCCGACGTCGGCCTTATTTAGCGCTTCGCCTATGGCGCAGCCGAG 377
          | |  |  |||| | |  | ||||   |  ||| |   | | | | |  | ||||  |
Db   1780 AACCATGACTTCTTCTCTCAAGTCGGAAATGGTTACCTTCTTGGCGAAGTTGAAGCCATG 1721

Qy    378 AACAACCATTTCTGATGCTGAGATTTGGGATA 409
          |  || ||  ||    ||  | ||| |  ||
Db   1720 CAAGACGATGACTATAACTACGGTTTAGACTA 1689


RESULT   8
Q03875
ID   Q03875 standard; DNA; 3095 BP.
AC   Q03875;
DT   24-AUG-1990  (first entry)
DE   Sequence encoding carboxylic terminal part of native GLURP.
KW   Plasmodium falciparum; antigen; malaria; vaccine; GLURP; ss.
OS   Plasmodium falciparum.
FH   Key             Location/Qualifiers
FT   cds             1. .2352
FT                   /*tag=  a
FT                   /product=GLURP
PN   WO9022811-A.
PD   22-MAR-1990.
PF   18-SEP-1989; 00218.
PR   03-MAR-1989; US-218885.
PR   03-MAR-1989; DK-005191.
PA   (STAT-) Statens Seruminst.
PI   Dziegiel M, Borre M, Jepsen S, Vuust J, Rieneck K, Wind A, Jakobsen PH;
DR   WPI; 90-115998/15.
DR   P-PSDB; R05804.
PT   Polypeptide(s) derived from Plasmidium falciparum antigen - used in
PT   vaccines and in production of antibodies, for diagnosis and
PT   therapy of malaria.
PS   Disclosure; Fig 7; 108pp; English.
CC   An open reading frame of 2349 bps extends from the 5' terminal end of the
CC   insert to a "TAA" stop codon. It is longest ORF found in the sequence.
CC   Sequence displays some of the characteristics of other malaria nucleic
CC   acid sequences : tandemly repeated motifs, high AT content and a
CC   corresponding preference for codons containing these bases, and a high
CC   content of codons for glutamate. Three major repetitive sequences are:
CC   one motif from bp 34 to bp 156 is repeated from bp 289 to bp 411; another
CC   motif from bp 477 to bp 521 is repeated tandemly twice from bp 522 to bp
CC   566 and from bp 567 to bp 611; a third motif from bp 1174 to bp 1233 is
CC   repeated tandemly 11 times.This last repetitious region consists of 360bp
CC   repeats differing only in 3 bases GAT coding for aspartate. This region
CC   is flanked to the 5' terminal od a degenerated 60 bp repeat. GC content
CC   of the coding part of the insert is on average 30%, and of the non-coding
CC   3' terminal 11% .
SQ   Sequence   3095 BP;    1443 A;    300 C;    491 G;    861 T;


  Query Match             3.2%;  Score 33.4;  DB 1;  Length 3095;
  Best Local Similarity   58.6%;  Pred. No. 1.7;
  Matches   58;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy    550 GGTAATACCGTTAAGCTAGATGTTACCTCACTCAAAAATGATTTAAACAGTTTAGTCAAT 609
          |||||||   || |  || || ||   | | |  |||    || |||||  | |   |||
Db   2903 GGTAATAAAATTCATATATATTTTTTTTAATTTTAAATATTTTAAAACAAATGATATAAT 2962

Qy    610 AAATATAATCAAATAAACAGTAATACCGTTTTATTTCCA 648
          || || |   || ||| ||||| |   | | || ||||
Db   2963 TAAAATTAAATAAGAAAAAGTAAGATTTTATAATGTCCA 3001


RESULT   9
T29248
ID   T29248 standard; DNA; 8133 BP.
AC   T29248;
DT   08-JUL-1996  (first entry)
DE   C. difficile toxin A gene.
KW   Toxin A; cytotoxin; enterotoxin; fusion protein; antitoxin; vaccine;
KW   immunogen; diarrhoea; diagnosis; therapy; ds.
OS   Clostridium difficile VPI strain 10463 (ATCC 10463).
PN   WO9612802-A1.
PD   02-MAY-1996.
PF   23-OCT-1995; U13737.
PR   24-OCT-1994; US-329154.
PR   16-MAR-1995; US-405496.
PR   14-APR-1995; US-422711.
PR   07-JUN-1995; US-480604.
PA   (OPHI-) OPHIDIAN PHARM INC.
PI   Firca JR,  Kink JA,  Padhye NV,  Stafford DC,  Thalley BS;
PI   Williams JA;
DR   WPI; 96-230603/23.
DR   P-PSDB; R95016.
PT   Fusion proteins comprising non-toxin protein and part of toxin -
PT   useful to form anti-toxins against Clostridium botulinum type A, and
PT   C. difficile type toxins, and to treat C. difficile intoxication,
PT   partic. diarrhoea
PS   Example 11; Page 290-302; 434pp; English.
CC   A Clostridium difficile VPI strain 10463 gene (T29248) codes for
CC   toxin A (R95016), a cytotoxin that plays a direct role in damaging
CC   gastrointestinal tissues and associated with diarrhoetic disease.
CC   The gene was cloned by PCR amplification (see T29251-54) from C.
CC   difficile genomic DNA.  Although high expression in transformed E.
```

```
PI   Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI   Rosen CA;
DR   WPI; 97-374922/35.
PT   Polynucleotide(s) and proteins derived from Staphylococcus aureus -
PT   stored on computer readable medium and used in the production of
PT   anti-S.aureus vaccines
PS   Claim 1; Page 1487-1488; 3271pp; English.
CC   This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC   of the invention. The DNA sequences are recorded on a computer readable
CC   medium, preferably selected from a floppy or hard disk, random access
CC   memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC   the S.aureus DNA sequences allows putative functions to be assigned so
CC   that protein-encoding or regulatory regions of commercial, therapeutic or
CC   industrial importance can be obtained. Specifically, sequences which are
CC   likely to encode antigens have been identified and these polypeptides can
CC   be used in a vaccine composition against S.aureus infection. The
CC   polypeptides can also be used in a kit for the immunodetection of
CC   S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC   including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC   skin and surgical wound infections, scalded skin syndrome, toxic shock
CC   syndrome, etc. Organisms transformed with the DNA sequences can be used
CC   for recombinant production of the polypeptides. The new DNA sequences
CC   (and their fragments) are useful as primers or probes for isolating
CC   homologues of any of the S.aureus DNA sequences contained on the
CC   computer readable medium.
SQ   Sequence   1523 BP;    446 A;    255 C;    198 G;    561 T;

  Query Match             3.4%;  Score 35.4;  DB 1;  Length 1523;
  Best Local Similarity   48.2%;  Pred. No. 0.26;
  Matches   96;  Conservative    1;  Mismatches  102;  Indels    0;  Gaps    0;

Qy     690 AGCGAGACAGTGGCTCAGTGAATTGAATTTACCGAATAGCTGCCTGAAATCTTATGGATC 749
           ||||| |  |:|| | | | |     |  |||  || | |     |||||     || |
Db    1202 AGCGAAAGCGWGGATGATTAACACACGTACACCAGATTGGTTATATAAATCACCGCGAAC 1143

Qy     750 CGGTTATGTCGTCACCGTTGATCTGACGCCATTACAAAAAATGGTTCAGGATATTGATGG 809
             ||  |    ||     ||   |||||||||| |||| || | |  |||||||  |
Db    1142 GACGTACTTACAAACATGGCATGGCACGCCATTAAAAAAGATTGGTTTGGATATTAGTAA 1083

Qy     810 TTTAGGCGCGCCGGGAAAAGACTCAAAACTCGAAATGGATAACGCCAAATATCAAGCCTG 869
           |      ||  |||| | |  |  ||    |   |||  |  ||| |  ||  |
Db    1082 CGTTAAAATGCTAGGAACAAATACTCAAAATTACCAAGATGGCTTTAAAAAAGAAAGCCA 1023

Qy     870 GCAGTCGGGTTTTAAAGCG 888
           | || || || |  || |
Db    1022 ACGGTGGGATTATCTAGTG 1004


RESULT   5
X23522/c
ID   X23522 standard; DNA; 13340 BP.
AC   X23522;
DT   17-JUN-1999  (first entry)
DE   O. longistaminata Xa21 gene family member D DNA.
KW   Xa21; receptor kinase-like protein; multigene family; RRK; rice;
KW   plant disease resistance; cassava; maize; tomato; Xanthomonas; ss.
OS   Oryza longistaminata.
PN   WO9909151-A2.
PD   25-FEB-1999.
PF   17-JUL-1998; U14841.
PR   13-AUG-1997; US-910386.
PA   (REGC ) UNIV CALIFORNIA.
PI   Hulbert SH, Richter T, Ronald PC, Song W, Szabo V,
PI   Wang G;
DR   WPI; 99-204431/17.
DR   P-PSDB; W93594.
PT   New RRK polynucleotides and nucleic acid constructs - used for
PT   generating transgenic plants resistant to Xanthomonas
PS   Claim 1; Page 31-35; 67pp; English.
CC   This invention describes a method for conferring disease resistance in
CC   plants. The invention describes the use of novel genes and proteins
CC   belonging to the Oryza longistaminata and Oryza sativa receptor
CC   kinase-like protein (RRK) Xa21 multigene family. Such genes from
CC   cassava, maize and tomato are also described. The genes and proteins can
CC   be used for enhancing resistance to Xanthomonas in a plant, preferably in
CC   rice or tomato.
SQ   Sequence   13340 BP;    3812 A;    2767 C;    2799 G;    3962 T;

  Query Match             3.3%;  Score 34.4;  DB 1;  Length 13340;
  Best Local Similarity   57.4%;  Pred. No. 1.8;
  Matches   62;  Conservative    0;  Mismatches   46;  Indels    0;  Gaps    0;

Qy     733 CTGAAATCTTATGGATCCGGTTATGTCGTCACCGTTGATCTGACGCCATTACAAAAAATG 792
           ||  | || | || ||   | |||    | |  || || |||| ||  |||| ||  |||
Db     572 CTTTACTCCTGTGCATTGTGATATTAAATTATAGTAGAACTGATGCATTTACTAATGATG 513

Qy     793 GTTCAGGATATTGATGGTTTAGGCGCGCCGGGAAAAGACTCAAAACTC 840
           ||| || | |  ||||| | | | ||    |||  | | |  | || |
Db     512 GTTGAGAAAAAAGATGGGTGATGAGCTTTTGGATGATAGTATAGACAC 465


RESULT   6
V16108
ID   V16108 standard; DNA; 109 BP.
AC   V16108;
DT   02-JUL-1998  (first entry)
DE   Synthetic BRSV Gsyn ORF fragment 3.
KW   BRSV; bovine respiratory syncytium virus; attenuated; vaccine; BHV;
KW   bovine herpes virus; Gsyn ORF; ds.
OS   Synthetic.
OS   Bovine respiratory syncytial virus.
PN   WO9804710-A2.
PD   05-FEB-1998.
PF   28-JUL-1997; IB1040.
PR   26-JUL-1996; US-687881.
PA   (ALKU ) AKZO NOBEL NV.
PI   Keil GM, Rijsewijk FAM;
DR   WPI; 98-130698/12.
PT   Synthetic bovine respiratory syncytium virus gene encoding natural
PT   protein - and attenuated recombinant bovine herpes virus containing
PT   this gene, useful in vaccines to protect against both viruses
PS   Example 2; Page 21; 68pp; English.
CC   The present sequence represents a DNA fragment of a novel synthetic
CC   bovine respiratory syncytium virus (BRSV) gene, designated BRSV Gsyn
CC   ORF, which has a non-natural nucleotide sequence encoding a naturally
CC   occurring protein. The present invention also describes live attenuated
CC   recombinant bovine herpes virus (BHV) that: (a) contains BRSV Gsyn ORF;
CC   (b) includes a heterologous gene fused to a class II membrane
CC   glycoprotein membrane anchor (A), or (c) carries a heterologous class II
CC   membrane glycoprotein. (A), or their derived particles, are used in
CC   vaccines to protect cattle against infection by both BHV and BRSV. They
CC   can also be used as carriers for other antigenic proteins from other
CC   pathogens. Vaccines are specifically given intranasally, but also
CC   contemplated are injection and oral delivery, also administration of
CC   DNA isolated from recombinant BHV. The usual dose is 103-1010 pfu. BHV
CC   infects only bovine species so is an ideal carrier for the BRSV, or
CC   other, genes. The BRSV Gsyn ORF gene, unlike the natural genes, is
CC   effectively expressed in recombinant BHV and the BRSV gG protein becomes
CC   incorporated into the viral envelope. This means that a booster dose
CC   will induce a second immune response against both viruses without
CC   infection being necessary.
SQ   Sequence   109 BP;    35 A;    39 C;    26 G;    9 T;

  Query Match             3.3%;  Score 34;  DB 1;  Length 109;
  Best Local Similarity   59.2%;  Pred. No. 0.17;
  Matches   58;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy     158 CTAAAATCCACCAGGCACAGCAGACGCTGCAGTCAACGCCACCGATCTCTGAAGAGAATA 217
           | |  |||| ||| | |||||| ||| | |    |||  ||| |||     ||| | |
Db       7 CGACGATCCAGCAGACGCAGCAGCCGCAGAACCACACGAGCCCGTTCTTCACGGAGCACA 66

Qy     218 ATGACGAGCGCACGCTGGCGCGCCAGCAGTTGACCAGC 255
           |  || || ||||||   || ||   |||   || | |
```

```
Db   2617 AATATATCGGCGATAGGTGACAGCTATCTGGGCGTTTATGAAAACGTTGTCGCAGTCTAT 2676

Qy    481 ACCGATTTTTATCAGGCCTTCAGTGATATTCTTTCCAAAATGGGAGGCTGGTTATTACCA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2677 ACCGATTTTTATCAGGCCTTCAGTGATATTCTTTCCAAAATGGGAGGCTGGTTATTACCA 2736

Qy    541 GGTAAGGACGGTAATACCGTTAAGCTAGATGTTACCTCACTCAAAAATGATTTAAACAGT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2737 GGTAAGGACGGTAATACCGTTAAGCTAGATGTTACCTCACTCAAAAATGATTTAAACAGT 2796

Qy    601 TTAGTCAATAAATATAATCAAATAAACAGTAATACCGTTTTATTTCCAGCGCAGTCAGGC 660
          ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   2797 TTAGTCAATAAATATAATCAAATAAACAGTAATAGCGTTTTATTTCCAGCGCAGTCAGGC 2856

Qy    661 AGCGGCGTTAAAGTAGCCACTGAAGCGGAAGCGAGACAGTGGCTCAGTGAATTGAATTTA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2857 AGCGGCGTTAAAGTAGCCACTGAAGCGGAAGCGAGACAGTGGCTCAGTGAATTGAATTTA 2916

Qy    721 CCGAATAGCTGCCTGAAATCTTATGGATCCGGTTATGTCGTCACCGTTGATCTGACGCCA 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2917 CCGAATAGCTGCCTGAAATCTTATGGATCCGGTTATGTCGTCACCGTTGATCTGACGCCA 2976

Qy    781 TTACAAAAAATGGTTCAGGATATTGATGGTTTAGGCGCGCCGGGAAAAGACTCAAAACTC 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2977 TTACAAAAAATGGTTCAGGATATTGATGGTTTAGGCGCGCCGGGAAAAGACTCAAAACTC 3036

Qy    841 GAAATGGATAACGCCAAATATCAAGCCTGGCAGTCGGGTTTTAAAGCGCAGGAAGAAAAT 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3037 GAAATGGATAACGCCAAATATCAAGCCTGGCAGTCGGGTTTTAAAGCGCAGGAAGAAAAT 3096

Qy    901 ATGAAAACCACATTACAGACGCTGACGCAAAAATATAGCAATGCCAATTCATTGTACGAC 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3097 ATGAAAACCACATTACAGACGCTGACGCAAAAATATAGCAATGCCAATTCATTGTACGAC 3156

Qy    961 AACCTGGTAAAAGTGCTGAGCAGTACGATAAGTAGCAGCCTGGAAACCGCCAAAAGCTTC 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3157 AACCTGGTAAAAGTGCTGAGCAGTACGATAAGTAGCAGCCTGGAAACCGCCAAAAGCTTC 3216

Qy   1021 CTGCAAGGATAA 1032
          ||||||||||||
Db   3217 CTGCAAGGATAA 3228
```

```
RESULT   3
X15189
ID   X15189 standard; DNA; 1236 BP.
AC   X15189;
DT   29-APR-1999  (first entry)
DE   DNA encoding a tarF enzyme of Staphylococcus aureus.
KW   CDP-glycerol:poly(glycerophosphate) glycerophospho transferase; tarF;
KW   wall teichoic acid synthesis; bacterial disease; otitis media;
KW   thyroiditis; infective endocarditis; secretory diarrhoea;
KW   splenic abscess; cerebral abscess; conjunctivitis; keratitis;
KW   toxic shock syndrome; impetigo; wound infection; septic arthritis;
KW   osteomyelitis; bacterial adhesion; matrix protein; body implant; ds.
OS   Staphylococcus aureus.
FH   Key             Location/Qualifiers
FT   CDS             1. .1233
FT                   /*tag= a
FT                   /product= tarF_enzyme
FT                   /note= "CDP-glycerol:poly(glycerophosphate)
FT                          glycerophospho transferase"
FT                   /transl_except= (pos: 922. .924, aa: Xaa)
FT                   /note= "Xaa=unknown"
PN   EP-891983-A1.
PD   20-JAN-1999.
PF   01-JUL-1998; 305252.
PR   18-JUL-1997; US-896343.
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PA   (SMIK ) SMITHKLINE BEECHAM PLC.
PI   Chalker AF, Fosberry AP, Ovari M, Van Horn S;
DR   WPI; 99-083518/08.
DR   P-PSDB; W97090.
PT   New Staphylococcus aureus CDP-glycerol:poly(glycerophosphate)
PT   glycerophospho transferase (tarF) enzyme polypeptide and
PT   polynucleotide - useful as diagnostic reagents and for prevention
PT   and treatment of Staphylococci infections
PS   Claim 4; Page 6; 26pp; English.
CC   The present sequence encodes a CDP-glycerol:poly(glycerophosphate)
CC   glycerophospho transferase (tarF) enzyme. The tarF enzyme is involved
CC   in the synthesis of wall teichoic acids. TarF polypeptides and
CC   polynucleotides are useful for diagnosing diseases related to over or
CC   underexpression of tarF protein by identifying mutations in the tarF
CC   gene, or determining tarF polypeptide or mRNA expression levels due to
CC   an infection of an organism with the tarF gene. tarF polypeptides are
CC   also useful for their antagonists and agonists. tarF polypeptides can
CC   induce an immune response to immunise and prevent disease. Diseases
CC   diagnosed, prevented or treated include bacterial diseases (especially
CC   Staphylococcus aureus), including infections of the upper and lower
CC   respiratory tract (e.g. otitis media, thyroiditis), cardiac (e.g.
CC   infective endocarditis), gastrointestinal (e.g. secretory diarrhoea,
CC   splenic abscess), CNS (e.g. cerebral abscess), eye (e.g. conjunctivitis,
CC   keratitis), kidney and urinary tract (e.g. toxic shock syndrome), skin
CC   (e.g. impetigo, wound infection), and bone and joint (e.g. septic
CC   arthritis, osteomyelitis). TarF polypeptides, polynucleotides and
CC   their (ant)agonists can prevent adhesion of bacteria to matrix proteins,
CC   and are useful for use on wounds and body implants to prevent bacterial
CC   infection.
SQ   Sequence   1236 BP;    453 A;    194 C;    227 G;    361 T;
```

```
  Query Match             3.5%;  Score 35.8;  DB 1;  Length 1236;
  Best Local Similarity   48.7%;  Pred. No. 0.17;
  Matches   97;  Conservative    0;  Mismatches  102;  Indels    0;  Gaps    0;

Qy    690 AGCGAGACAGTGGCTCAGTGAATTGAATTTACCGAATAGCTGCCTGAAATCTTATGGATC 749
          ||||| |  |||| | | | |      |  |||  || | |     |||||     || |
Db    339 AGCGAAAGCGTGGATGATTAACACACGTACACCAGATTGGTTATATAAATCACCGCGAAC 398

Qy    750 CGGTTATGTCGTCACCGTTGATCTGACGCCATTACAAAAAATGGTTCAGGATATTGATGG 809
            ||  |    ||     ||   ||||||||| |||| || | |  |||||||  |
Db    399 GACGTACTTACAAACATGGCATGGCACGCCATTAAAAAAGATTGGTTTGGATATTAGTAA 458

Qy    810 TTTAGGCGCGCCGGGAAAAGACTCAAAACTCGAAATGGATAACGCCAAATATCAAGCCTG 869
          |      ||  |||| | |  |  ||    |   |||  |   ||| |  ||  |
Db    459 CGTTAAAATGCTAGGAACAAATACTCAAAATTACCAAGATGGCTTTAAAAAAGAAAGCCA 518

Qy    870 GCAGTCGGGTTTTAAAGCG 888
          | || || || |  || |
Db    519 ACGGTGGGATTATCTAGTG 537
```

```
RESULT   4
V74861/c
ID   V74861 standard; DNA; 1523 BP.
AC   V74861;
DT   16-MAR-1999  (first entry)
DE   Staphylococcus aureus contig SEQ ID #550.
KW   Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW   cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW   skin infection; surgical wound infection; scalded skin syndrome;
KW   toxic shock syndrome; ds.
OS   Staphylococcus aureus.
FH   Key             Location/Qualifiers
FT   misc_feature    361. .420
FT                   /*tag= a
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
FT                          are included to maintain the nucleotide numbering
FT                          given in the specification for this DNA sequence"
PN   EP-786519-A2.
PD   30-JUL-1997.
PF   07-JAN-1997; 100117.
PR   05-JAN-1996; US-009861.
PA   (HUMA-) HUMAN GENOME SCI INC.
```

```
Db    181 ACGCTGCAGTCAACGCCACCGATCTCTGAAGAGAATAATGACGAGCGCACGCTGGCGCGC 240

Qy    241 CAGCAGTTGACCAGCAGCCTGAATGCGCTGGCGAAGTCCGGCGTGTCATTATCCGCAGAA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CAGCAGTTGACCAGCAGCCTGAATGCGCTGGCGAAGTCCGGCGTGTCATTATCCGCAGAA 300

Qy    301 CAAAATGAGAACCTGCGGAGCGCGTTTTCTGCGCCGACGTCGGCCTTATTTAGCGCTTCG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CAAAATGAGAACCTGCGGAGCGCGTTTTCTGCGCCGACGTCGGCCTTATTTAGCGCTTCG 360

Qy    361 CCTATGGCGCAGCCGAGAACAACCATTTCTGATGCTGAGATTTGGGATATGGTTTCCCAA 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CCTATGGCGCAGCCGAGAACAACCATTTCTGATGCTGAGATTTGGGATATGGTTTCCCAA 420

Qy    421 AATATATCGGCGATAGGTGACAGCTATCTGGGCGTTTATGAAAACGTTGTCGCAGTCTAT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AATATATCGGCGATAGGTGACAGCTATCTGGGCGTTTATGAAAACGTTGTCGCAGTCTAT 480

Qy    481 ACCGATTTTTATCAGGCCTTCAGTGATATTCTTTCCAAAATGGGAGGCTGGTTATTACCA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 ACCGATTTTTATCAGGCCTTCAGTGATATTCTTTCCAAAATGGGAGGCTGGTTATTACCA 540

Qy    541 GGTAAGGACGGTAATACCGTTAAGCTAGATGTTACCTCACTCAAAAATGATTTAAACAGT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 GGTAAGGACGGTAATACCGTTAAGCTAGATGTTACCTCACTCAAAAATGATTTAAACAGT 600

Qy    601 TTAGTCAATAAATATAATCAAATAAACAGTAATACCGTTTTATTTCCAGCGCAGTCAGGC 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 TTAGTCAATAAATATAATCAAATAAACAGTAATACCGTTTTATTTCCAGCGCAGTCAGGC 660

Qy    661 AGCGGCGTTAAAGTAGCCACTGAAGCGGAAGCGAGACAGTGGCTCAGTGAATTGAATTTA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 AGCGGCGTTAAAGTAGCCACTGAAGCGGAAGCGAGACAGTGGCTCAGTGAATTGAATTTA 720

Qy    721 CCGAATAGCTGCCTGAAATCTTATGGATCCGGTTATGTCGTCACCGTTGATCTGACGCCA 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 CCGAATAGCTGCCTGAAATCTTATGGATCCGGTTATGTCGTCACCGTTGATCTGACGCCA 780

Qy    781 TTACAAAAAATGGTTCAGGATATTGATGGTTTAGGCGCGCCGGGAAAAGACTCAAAACTC 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 TTACAAAAAATGGTTCAGGATATTGATGGTTTAGGCGCGCCGGGAAAAGACTCAAAACTC 840

Qy    841 GAAATGGATAACGCCAAATATCAAGCCTGGCAGTCGGGTTTTAAAGCGCAGGAAGAAAAT 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 GAAATGGATAACGCCAAATATCAAGCCTGGCAGTCGGGTTTTAAAGCGCAGGAAGAAAAT 900

Qy    901 ATGAAAACCACATTACAGACGCTGACGCAAAAATATAGCAATGCCAATTCATTGTACGAC 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 ATGAAAACCACATTACAGACGCTGACGCAAAAATATAGCAATGCCAATTCATTGTACGAC 960

Qy    961 AACCTGGTAAAAGTGCTGAGCAGTACGATAAGTAGCAGCCTGGAAACCGCCAAAAGCTTC 1020
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    961 AACCTGGTAAAAGTGCTGAGCACTACGATAAGTAGCAGCCTGGAAACCGCCAAAAGCTTC 1020

Qy   1021 CTGCAAGGATAA 1032
          ||||||||||||
Db   1021 CTGCAAGGATAA 1032


RESULT  2
T67035
ID   T67035 standard; DNA; 3441 BP.
AC   T67035;
DT   04-AUG-1997  (first entry)
DE   Salmonella virulence factor SspBCDA DNA region.
KW   Salmonella secreted protein; SspB; SspC; SspD; SspA;
KW   bacterial-mediated endocytosis; diagnosis; therapy; vaccine;
KW   attenuation; virulence; ss.
OS   Salmonella typhimurium.
FH   Key             Location/Qualifiers
FT   cds             1. .870
FT                   /*tag=  a
FT                   /product= SspB (incomplete)
FT   cds             898. .2127
FT                   /*tag=  b
FT                   /product= SspC
FT   cds             2198. .3228
FT                   /*tag=  c
FT                   /product= SspD
FT   cds             3247. .3441
FT                   /*tag=  d
FT                   /product= SspA (incomplete)
FT
PN   WO9718225-A1.
PD   22-MAY-1997.
PF   14-NOV-1996; U18504.
PR   14-NOV-1995; US-006733.
PA   (GEHO ) GEN HOSPITAL CORP.
PI   Miller SI;
DR   WPI; 97-289217/26.
DR   P-PSDB; W15268-71.
PT   New isolated Salmonella secreted proteins and related genes - used
PT   to develop products for the detection, treatment or prevention of
PT   Salmonella infections
PS   Claim 11; Fig 28; 95pp; English.
CC   A DNA molecule (T67035) comprises the SspBCDA locus (truncated at
CC   the 5' and 3' ends) including the partial sspB, sspC, sspD and
CC   partial sspA sequences (see also T67028-31) coding for Salmonella
CC   typhimurium secreted proteins SspB, SspC, SspD and SspA (W15268-
CC   70).  To identify the ssp genes, S. typhimurium mutants were
CC   investigated in invasion assays on HEp-2 epithelial cells.  2
CC   Ssp mutants with transposon insertions located between spaT and
CC   prgH (see also T67036) were identified.  One mutant lacked Ssp of
CC   87, 42 and 36 kDa.  The invasiveness of this mutant implicated the
CC   42 and 36 kDa Ssp in invasion.  DNA sequencing of this region
CC   identified the 4 ssp (partial) genes.  The ssp genes can be used in
CC   methods for diagnosing Salmonella infection and in methods of
CC   targetting an antigen to epithelial cells and of inducing a
CC   cytotoxic T cell immune response in a mammal.  Ssp gene mutation
CC   can be used to create attenuated Salmonella strains useful as live
CC   vaccines.
SQ   Sequence   3441 BP;    978 A;    761 C;    925 G;    777 T;


  Query Match             98.7%;  Score 1018.4;  DB 1;  Length 3441;
  Best Local Similarity   99.8%;  Pred. No. 0;
  Matches 1030;  Conservative    0;  Mismatches    1;  Indels    1;  Gaps    1;

Qy      1 ATGCTTAATATTCAAAATTATTCCGCTTCTCCTCATCCGGGGATCGTTGCCGAACGGCCG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2198 ATGCTTAATATTCAAAATTATTCCGCTTCTCCTCATCCGGGGATCGTTGCCGAACGGCCG 2257

Qy     61 CAGACTCCCTCGGCGAGCGAGCACGTCGAGACTGCCGTGGTACCGTCTACCACAGAACAT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2258 CAGACTCCCTCGGCGAGCGAGCACGTCGAGACTGCCGTGGTACCGTCTACCACAGAACAT 2317

Qy    121 CGCGGTACAGATATCATTTCATTATCGCAGGCGGCTACTAAAATCCACCAGGCACAGCAG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2318 CGCGGTACAGATATCATTTCATTATCGCAGGCGGCTACTAAAATCCACCAGGCACAGCAG 2377

Qy    181 ACGCTGCAGTCAACGCCACCGATCTCTGAAGAGAATAATGACGAGCGCACGCTGGCGCGC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2378 ACGCTGCAGTCAACGCCACCGATCTCTGAAGAGAATAATGACGAGCGCACGCTGGCGCGC 2437

Qy    241 CAGCAGTTGACCAGCAGCCTGAATGCGCTGGCGAAGTCCGGCGTGTCATTATCCGCAGAA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2438 CAGCAGTTGACCAGCAGCCTGAATGCGCTGGCGAAGTCCGGCGTGTCATTATCCGCAGAA 2497

Qy    301 CAAAATGAGAACCTGCGGAGCGCGTTTTCTGCGCCGACGTCGGCCTTATTTAGCGCTTCG 360
          |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db   2498 CAAAATGAGAACCTGCGGAGCGCGTTTTCTGCGCCGACGTCGGCCTTA-TTAGCGCTTCG 2556

Qy    361 CCTATGGCGCAGCCGAGAACAACCATTTCTGATGCTGAGATTTGGGATATGGTTTCCCAA 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2557 CCTATGGCGCAGCCGAGAACAACCATTTCTGATGCTGAGATTTGGGATATGGTTTCCCAA 2616

Qy    421 AATATATCGGCGATAGGTGACAGCTATCTGGGCGTTTATGAAAACGTTGTCGCAGTCTAT 480
```

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 1999, 17:58:46 ; Search time 223.92 Seconds
(without alignments)
1153.082 Million cell updates/sec

Title: US-09-068-804-3
Perfect score: 1032
Sequence: 1 ATGCTTAATATTCAAAATTA.....AAAGCTTCCTGCAAGGATAA 1032

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1030.4	99.8	1032	1	T67030	Salmonella secrete
2	1018.4	98.7	3441	1	T67035	Salmonella virulen
3	35.8	3.5	1236	1	X15189	DNA encoding a tar
c 4	35.4	3.4	1523	1	V74861	Staphylococcus aur
c 5	34.4	3.3	13340	1	X23522	O. longistaminata
6	34	3.3	109	1	V16108	Synthetic BRSV Gsy
c 7	33.6	3.3	3189	1	V65242	DNA encoding three
8	33.4	3.2	3095	1	Q03875	Sequence encoding
9	33.4	3.2	8133	1	T29248	C. difficile toxin
10	33.4	3.2	8133	1	V30560	Clostridium diffic
11	33.2	3.2	1560	1	X00225	Clostridium chauvo
c 12	32.8	3.2	3902	1	V52345	Streptococcus pneu
13	32.4	3.1	5132	1	V74474	Staphylococcus aur
c 14	32.2	3.1	529	1	Q39758	Expressed Sequence
c 15	32.2	3.1	529	1	Q59170	Human brain Expres
16	31.8	3.1	110000	1	V21209_07	Continuation (8 of
17	31.6	3.1	5437	1	Q13369	Fusaric acid resis
c 18	31.6	3.1	6022	1	V74403	Staphylococcus aur
19	31.6	3.1	1471	1	X07096	Staphylococcus aur
20	31.6	3.1	1471	1	X07097	Staphylococcus aur
c 21	31	3.0	3754	1	V74869	Staphylococcus aur
22	31	3.0	3958	1	V74480	Staphylococcus aur
23	31	3.0	110000	1	X20248_03	Continuation (4 of
c 24	30.8	3.0	1524	1	Q55760	Delta-9-desaturase
c 25	30.8	3.0	2292	1	V84357	Streptococcus pneu
26	30.6	3.0	814	1	Q14191	Promoter from pXXL
27	30.6	3.0	110000	1	V21209_09	Continuation (10 o
28	30.6	3.0	1820	1	V63896	Bacillus sp. phyta
c 29	30.4	2.9	669	1	Q06197	Encodes Porcine Br
c 30	30.4	2.9	1741	1	T11674	DNA encoding a pro
c 31	30.4	2.9	1738	1	T86130	Arabidopsis protop
c 32	30.4	2.9	1662	1	T95392	Arabidopsis thalia
c 33	30.4	2.9	1738	1	V04305	Arabidopsis thalia
c 34	30.4	2.9	1738	1	V07252	Arabidopsis cDNA e
c 35	30.4	2.9	20986	1	V52187	Streptococcus pneu
c 36	30.4	2.9	21126	1	X12960	Enterococcus faeca
37	30.2	2.9	7462	1	Q35153	Jaagsiekte retrovi
c 38	30.2	2.9	6511	1	Q95493	Human Cdn-2 DNA. N
c 39	30.2	2.9	1770	1	T14228	Human neural cell
40	30.2	2.9	1776	1	V59559	Human secreted pro
c 41	30.2	2.9	3265	1	V99563	Maize lysine ketog
42	30	2.9	11715	1	N81564	Genomic sequence e
43	30	2.9	11724	1	N70102	Complete genomic s
44	30	2.9	4814	1	Q32956	PPI gene in phage
45	30	2.9	15223	1	T63430	Respiratory syncyt

ALIGNMENTS

RESULT 1
T67030
ID T67030 standard; DNA; 1032 BP.
AC T67030;
DT 04-AUG-1997 (first entry)
DE Salmonella secreted protein SspD DNA.
KW Salmonella secreted protein; SspD; bacterial-mediated endocytosis;
KW diagnosis; therapy; vaccine; attenuation; virulence; ss.
OS Salmonella typhimurium.
FH Key Location/Qualifiers
FT cds 1. .1032
FT /*tag= a
FT /product= SspD
FT /note= "the SspD amino acid sequence deduced
FT from the coding sequence does not fully
FT correspond to the SspD sequence given in
FT Fig 14"
PN WO9718225-A1.
PD 22-MAY-1997.
PF 14-NOV-1996; U18504.
PR 14-NOV-1995; US-006733.
PA (GEHO) GEN HOSPITAL CORP.
PI Miller SI;
DR WPI; 97-289217/26.
DR P-PSDB; W15270.
PT New isolated Salmonella secreted proteins and related genes - used
PT to develop products for the detection, treatment or prevention of
PT Salmonella infections
PS Claim 7; Fig 21; 95pp; English.
CC A DNA molecule (T67030) codes for Salmonella secreted protein SspD
CC (W15270), a virulence factor involved in bacterial mediated
CC endocytosis, which is able to translocate into human epithelial
CC cells. To identify Ssp genes, S. typhimurium mutants were
CC investigated in invasion assays on HEp-2 epithelial cells. 2
CC Ssp mutants with transposon insertions located between spaT and
CC prgH were identified. Of these, one mutant lacked Ssp of 87,
CC 42 and 36 kDa. The invasiveness of this mutant implicated the
CC 42 and 36 kDa Ssp in invasion. DNA sequencing of this region (see
CC also T67035) identified 2 complete and 2 partial open reading
CC frames, designated sspB, sspC (T67029), sspD (T67030) and sspA
CC (T67031). These nucleic acids can be used in methods for diagnosing
CC Salmonella infection and in methods of targetting an antigen to
CC epithelial cells and of inducing a cytotoxic T cell immune response
CC in a mammal. Ssp gene mutation can be used to create attenuated
CC Salmonella strains useful as live vaccines.
SQ Sequence 1032 BP; 304 A; 242 C; 246 G; 240 T;

Query Match 99.8%; Score 1030.4; DB 1; Length 1032;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1031; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      1 ATGCTTAATATTCAAAATTATTCCGCTTCTCCTCATCCGGGGATCGTTGCCGAACGGCCG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGCTTAATATTCAAAATTATTCCGCTTCTCCTCATCCGGGGATCGTTGCCGAACGGCCG 60

Qy     61 CAGACTCCCTCGGCGAGCGAGCACGTCGAGACTGCCGTGGTACCGTCTACCACAGAACAT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CAGACTCCCTCGGCGAGCGAGCACGTCGAGACTGCCGTGGTACCGTCTACCACAGAACAT 120

Qy    121 CGCGGTACAGATATCATTTCATTATCGCAGGCGGCTACTAAAATCCACCAGGCACAGCAG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CGCGGTACAGATATCATTTCATTATCGCAGGCGGCTACTAAAATCCACCAGGCACAGCAG 180

Qy    181 ACGCTGCAGTCAACGCCACCGATCTCTGAAGAGAATAATGACGAGCGCACGCTGGCGCGC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

This Page Blank (uspto)

Search completed: October 5, 1999, 18:01:01
Job time: 11501 sec

```
Qy      63 GACTCCCTCGGCGAGCGAGCACGTCGAGACTGCCGTGGTACCGTCTACCACAGAACATCG 122
           |||| |  |  ||   ||  |  ||     | |   ||| |||| | ||||  |   |
Db     746 TTCTCCATGAGATAGATGGCTGGCAGAAGTGCCAGCCTTACTGTCTGCAACAGTTCTGGG 687

Qy     123 CGGTACAGATATCATTTCATTATCGCA 149
           ||||   ||| |  ||| | |   ||
Db     686 AGGTAGCAATAGCGCTTCTTCAGGTCA 660


RESULT 14
US-08-756-506-1
; Sequence 1, Application US/08756506
; Patent No. 5905185
;  GENERAL INFORMATION:
;    APPLICANT:  Garner, Ian
;    APPLICANT:  Cottingham, Ian R.
;    APPLICANT:  Temperley, Simon M.
;    APPLICANT:  Foster, Donald C.
;    APPLICANT:  Sprecher, Cindy A.
;    APPLICANT:  Prunkard, Donna E.
;    TITLE OF INVENTION:  PROTEIN C PRODUCTION IN TRANSGENIC
;    TITLE OF INVENTION:  ANIMALS
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ZymoGenetics, Inc.
;      STREET:  1201 Eastlake Avenue East
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98102
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/756,506
;      FILING DATE:
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sawislak, Deborah A
;      REGISTRATION NUMBER:  37,438
;      REFERENCE/DOCKET NUMBER:  95-28
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-442-6672
;      TELEFAX:  206-442-6678
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11725 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(3520..3630, 5093..5117, 5210..5347, 5450
;      LOCATION:  ..5584, 8253..8395, 9269..9386, 10516..11102)
US-08-756-506-1


  Query Match             2.9%;  Score 30;  DB 4;  Length 11725;
  Best Local Similarity   54.5%;  Pred. No. 17;
  Matches   60;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy     521 TGGGAGGCTGGTTATTACCAGGTAAGGACGGTAATACCGTTAAGCTAGATGTTACCTCAC 580
           ||||||  ||    |  || ||  |  || || | ||| |    |||  | | |    |
Db    6797 TGGGAGTTTGAGACTAGCCTGGGCAACACAGTGAGACCCTGTCTCTATTTTTAAAAAAAG 6856

Qy     581 TCAAAAATGATTTAAACAGTTTAGTCAATAAATATAATCAAATAAACAGT 630
           | |||||  ||| |||| | || | |   ||  |  ||  || || | | |
Db    6857 TAAAAAAAGATCTAAAAATTTAACTTTTTATTTTGAAATAATTAGATATT 6906
```

```
RESULT 15
US-08-920-812-10/c
; Sequence 10, Application US/08920812
; Patent No. 5763188
;  GENERAL INFORMATION:
;    APPLICANT:  Ohno, Tsuneya
;    APPLICANT:  Matsuhisa, Akio
;    APPLICANT:  Uehara, Hirotsugu
;    APPLICANT:  Eda, Soji
;    TITLE OF INVENTION:  Probe for Diagnosing Infectious Disease
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/920,812
;      FILING DATE:  29-AUG-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/362,577
;      FILING DATE:  27-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rin-Laures, Li-Hsien
;      REGISTRATION NUMBER:  33,547
;      REFERENCE/DOCKET NUMBER:  19036/32420
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3719 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Enterococcus faecalis
;      STRAIN:  Clinical Isolate S2-3
US-08-920-812-10


  Query Match             2.9%;  Score 29.8;  DB 3;  Length 3719;
  Best Local Similarity   47.2%;  Pred. No. 9.7;
  Matches   91;  Conservative    0;  Mismatches  102;  Indels    0;  Gaps    0;

Qy     551 GTAATACCGTTAAGCTAGATGTTACCTCACTCAAAAATGATTTAAACAGTTTAGTCAATA 610
           |||  |||   | | ||| | ||  | || |   ||||  |||||  | ||  | || |
Db    1537 GTATAACCAAAAGGATAGCTTTTTACACAATGCTAAATTTATTAAATGGCTTCCTAAAGA 1478

Qy     611 AATATAATCAAATAAACAGTAATACCGTTTTATTTCCAGCGCAGTCAGGCAGCGGCGTTA 670
           ||   | ||||| ||||  |   |      ||||  ||||     |   |   |  | |
Db    1477 AAGTCATTCAAAAAAACGTTCGCAAGAGGATATTCACAGCCATTACGATCTAGGAAATGA 1418

Qy     671 AAGTAGCCACTGAAGCGGAAGCGAGACAGTGGCTCAGTGAATTGAATTTACCGAATAGCT 730
           |  | |   | |       | | ||| || |  | | |  ||  | |    ||  ||
Db    1417 TTTTTACAAAAAATGGCTTGACCAAACAATGACGTATTCATGTGCTTATTTTAAAACGCC 1358

Qy     731 GCCTGAAATCTTA 743
           ||  |  |||
Db    1357 TGAAGATACATTA 1345
```

```
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  100...492
;      OTHER INFORMATION:
US-08-850-910A-40
```

```
  Query Match              2.9%;  Score 30.4;  DB 4;  Length 707;
  Best Local Similarity   50.0%;  Pred. No. 2.1;
  Matches   76;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy     608 ATAAATATAATCAAATAAACAGTAATACCGTTTTATTTCCAGCGCAGTCAGGCAGCGGCG 667
           | | | ||||  | ||||| |  || |    ||  ||| || | ||   || || ||||
Db     593 AAATAAATAAATACATAAATAAAAACAGGAGTTGCTTTGAAGGGGAGCAGGGGAGGGGCG 534

Qy     668 TTAAAGTAGCCACTGAAGCGGAAGCGAGACAGTGGCTCAGTGAATTGAATTTACCGAATA 727
           ||  || |||       || ||  |  || |||    ||   | |      |   | | |
Db     533 TTGGAGAAGCGGACACAGAGGTTGTCAGCCAGGACTTCTCAGTACCTCCTGAGCACATTG 474

Qy     728 GCTGCCTGAAATCTTATGGATCCGGTTATGTC 759
               || |    || | ||| ||| |   |||
Db     473 CAGCCCAGGCCGCTGAGGGAGCCGATCCGGTC 442
```

```
RESULT  12
US-08-317-305-1/c
; Sequence 1, Application US/08317305
; Patent No. 5863744
;  GENERAL INFORMATION:
;    APPLICANT:  Avraham, Shalom
;    APPLICANT:  Avraham, Hava
;    APPLICANT:  Groopman, Jerome E.
;    TITLE OF INVENTION:  NOVEL PROTEIN MARKER AND DNA ENCODING
;    TITLE OF INVENTION:  SAME
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Weingarten, Schurgin, Gagnebin & Hayes
;      STREET:  Ten Post Office Square
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  US
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/317,305
;      FILING DATE:  03-OCT-1994
;      CLASSIFICATION:  436
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Williams Ph.D., Kathleen A.
;      REGISTRATION NUMBER:  34,380
;      REFERENCE/DOCKET NUMBER:  DH-001XX
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 542-2290
;      TELEFAX:  (617) 451-0313
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1770 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1767
US-08-317-305-1
```

```
  Query Match              2.9%;  Score 30.2;  DB 3;  Length 1770;
  Best Local Similarity   50.3%;  Pred. No. 4.4;
  Matches   74;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy       3 GCTTAATATTCAAAATTATTCCGCTTCTCCTCATCCGGGGATCGTTGCCGAACGGCCGCA 62
           || |  |   |||  |  || | |||   ||  |  | |    ||| |   | |||| ||
Db     806 GCCTCTTCCACAATTTCCTTACTCTTTCTCTGCTTGGTGATGAGTTCCTCCATGGCCACA 747

Qy      63 GACTCCCTCGGCGAGCGAGCACGTCGAGACTGCCGTGGTACCGTCTACCACAGAACATCG 122
             |||| |  |  ||   ||  |  ||      | |   ||| |||| | |||| |   |
Db     746 TTCTCCATGAGATAGATGGCTGGCAGAAGTGCCAGCCTTACTGTCTGCAACAGTTCTGGG 687

Qy     123 CGGTACAGATATCATTTCATTATCGCA 149
             ||||   ||| |  ||| | |   ||
Db     686 AGGTAGCAATAGCGCTTCTTCAGGTCA 660
```

```
RESULT  13
PCT-US95-12508-1/c
; Sequence 1, Application PC/TUS9512508
;  GENERAL INFORMATION:
;    APPLICANT:  Avraham, Shalom
;    APPLICANT:  Avraham, Hava
;    APPLICANT:  Groopman, Jerome E.
;    TITLE OF INVENTION:  NOVEL PROTEIN MARKER AND DNA ENCODING SAME
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lahive & Cockfield
;      STREET:  60 State Street, Suite 510
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  US
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/12508
;      FILING DATE:  29-SEP-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/317,305
;      FILING DATE:  03-OCT-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Silveri, Jean M.
;      REGISTRATION NUMBER:  39,030
;      REFERENCE/DOCKET NUMBER:  NER-259PC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 227-7400
;      TELEFAX:  (617) 227-5941
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1770 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1767
PCT-US95-12508-1
```

```
  Query Match              2.9%;  Score 30.2;  DB 5;  Length 1770;
  Best Local Similarity   50.3%;  Pred. No. 4.4;
  Matches   74;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy       3 GCTTAATATTCAAAATTATTCCGCTTCTCCTCATCCGGGGATCGTTGCCGAACGGCCGCA 62
           || |  |   |||  |  || | |||   ||  |  | |    ||| |   | |||| ||
Db     806 GCCTCTTCCACAATTTCCTTACTCTTTCTCTGCTTGGTGATGAGTTCCTCCATGGCCACA 747
```

```
Qy    380 CAACCATTTCTGATGCTGAGATTTGGGATATGGTTTCCCAAAATATATCGGCGATAGGTG 439
          |||| |  | |||  || ||||| || | |  |    | |       | | ||  || | |
Db    474 CAACAAGATTTGAAACTTAGATTGGGTAGAGAGCACACTACTTGTGACCAGCTCTATGGG 415

Qy    440 ACAGCTATCTGGGCGTTTATGAAAACGTTGTCGCAGTCTATACCGATTTTTATCAGGCCT 499
          |  | ||  | |       | | | ||  |  ||||   |||||||  |||||  |  |   |
Db    414 ATTGGTAGGTAGCATCACAGGTACACCATTCCGCACAATATACCGCTTTTTCTGTGAAAT 355

Qy    500 TCAGTGATATTCTTTC 515
          |      | || ||||
Db    354 TGGAAATTGTTGTTTC 339


RESULT 10
US-08-808-931-3/c
; Sequence 3, Application US/08808931
; Patent No. 5939602
;  GENERAL INFORMATION:
;    APPLICANT:  Volrath, Sandra
;    APPLICANT:  Johnson, Marie
;    APPLICANT:  Potter, Sharon
;    APPLICANT:  Ward, Eric
;    APPLICANT:  Heifetz, Peter
;    TITLE OF INVENTION:  DNA Molecules Encoding Plant
;    TITLE OF INVENTION:  Protoporphyrinogen Oxidase and Inhibitor-Resistant Mutants
;    TITLE OF INVENTION:  Thereof
;    NUMBER OF SEQUENCES:  35
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5939602artis Corporation
;      STREET:  520 White Plains Road, P.O. Box 2005
;      CITY:  Tarrytown
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10591-9005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/808,931
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/012,705
;      FILING DATE:  28-FEB-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/013,612
;      FILING DATE:  28-FEB-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/020,003
;      FILING DATE:  21-JUN-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  CGC 1847
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (919) 541-8587
;      TELEFAX:  (919) 541-8689
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1738 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Arabidopsis thaliana
;    IMMEDIATE SOURCE:
;      CLONE:  pWDC-1 (NRRL B-21237)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  70..1596
;      OTHER INFORMATION:  /product= "Arabidopsis protox-2"
US-08-808-931-3


  Query Match             2.9%;  Score 30.4;  DB 4;  Length 1738;
  Best Local Similarity   51.5%;  Pred. No. 3.8;
  Matches   70;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy    380 CAACCATTTCTGATGCTGAGATTTGGGATATGGTTTCCCAAAATATATCGGCGATAGGTG 439
          |||| |  | |||  || ||||| || | |  |    | |       | | ||  || | |
Db    474 CAACAAGATTTGAAACTTAGATTGGGTAGAGAGCACACTACTTGTGACCAGCTCTATGGG 415

Qy    440 ACAGCTATCTGGGCGTTTATGAAAACGTTGTCGCAGTCTATACCGATTTTTATCAGGCCT 499
          |  | ||  | |       | | | ||  |  ||||   |||||||  |||||  |  |   |
Db    414 ATTGGTAGGTAGCATCACAGGTACACCATTCCGCACAATATACCGCTTTTTCTGTGAAAT 355

Qy    500 TCAGTGATATTCTTTC 515
          |      | || ||||
Db    354 TGGAAATTGTTGTTTC 339


RESULT 11
US-08-850-910A-40/c
; Sequence 40, Application US/08850910A
; Patent No. 5948761
;  GENERAL INFORMATION:
;    APPLICANT:  SEILHAMER, J.J.
;    APPLICANT:  LEWICKI, J.
;    APPLICANT:  SCARBOROUGH, R.M.
;    TITLE OF INVENTION:  RECOMBINANT TECHNIQUES FOR
;    TITLE OF INVENTION:  PRODUCTION OF BRAIN NATRIUETIC PEPTIDE
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER, LLP
;      STREET:  2000 Pennsylvania Avenue, NW, Suite 5500
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20006-1888
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows
;      SOFTWARE:  FastSEQ for Windows Version 2.0b
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/850,910A
;      FILING DATE:  05-MAY-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/477,226
;      FILING DATE:  08-FEB-1990
;      APPLICATION NUMBER:  07/299,880
;      FILING DATE:  19-JAN-1989
;      APPLICATION NUMBER:  07/206,470
;      FILING DATE:  14-JUN-1988
;      APPLICATION NUMBER:  07/200,383
;      FILING DATE:  31-MAY-1988
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Murashige, Kate H
;      REGISTRATION NUMBER:  29,959
;      REFERENCE/DOCKET NUMBER:  219002025212
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-887-1500
;      TELEFAX:  202-822-0168
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  40:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  707 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
```

```
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  609-520-3214
;      TELEFAX:  609-520-3259
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2292 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-889-711-1


  Query Match             3.0%;  Score 30.8;  DB 3;  Length 2292;
  Best Local Similarity   61.0%;  Pred. No. 3.3;
  Matches   50;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy     556 ACCGTTAAGCTAGATGTTACCTCACTCAAAAATGATTTAAACAGTTTAGTCAATAAATAT 615
           ||  ||| ||| | || ||   | |||||||| | || |   || |  | | |  ||  |
Db    1245 ACGTTTAGGCTCGTTGATAATACGCTCAAAACTAATATTGTCACTCAAATTAGCAATAAA 1186

Qy     616 AATCAAATAAACAGTAATACCG 637
           || | |||| || ||||| ||
Db    1185 GATGAGATAAGCAATAATATCG 1164


RESULT   8
PCT-US95-04228-32
; Sequence 32, Application PC/TUS9504228
;  GENERAL INFORMATION:
;    APPLICANT:  Genentech, Inc.
;    APPLICANT:  Bennett, Brian D.
;    APPLICANT:  Goeddel, David
;    APPLICANT:  Lee, James M.
;    APPLICANT:  Matthews, William
;    APPLICANT:  Tsai, Siao Ping
;    APPLICANT:  Wood, William I.
;    TITLE OF INVENTION:  PROTEIN TYROSINE KINASE AGONIST ANTIBODIES
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/04228
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/222616
;      FILING DATE:  04-APR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Wendy M. Lee
;      REGISTRATION NUMBER:  00,000
;      REFERENCE/DOCKET NUMBER:  821P3PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-1994
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  32:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4425 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
PCT-US95-04228-32


  Query Match             3.0%;  Score 30.8;  DB 5;  Length 4425;
  Best Local Similarity   50.0%;  Pred. No. 5.1;
  Matches   77;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy     316 CGGAGCGCGTTTTCTGCGCCGACGTCGGCCTTATTTAGCGCTTCGCCTATGGCGCAGCCG 375
           |||  | ||  | ||  ||| | ||  |||||   ||    |  | ||| || | |  ||
Db     213 CGGTCCTCGAGTCCTCCGCGGTCGGTGGCCTCTGTTCCTGTCGCTCCTGTGCCCCCACCA 272

Qy     376 AGAACAACCATTTCTGATGCTGAGATTTGGGATATGGTTTCCCAAAATATATCGGCGATA 435
           |  |   |  | | |   ||| | |  ||||   ||| | |||     |   ||   |
Db     273 CGCTCTGACGCTCCCGTGTCTGCGGTCCGGGATGACGTTCCACAACGACGACGTGCTCCA 332

Qy     436 GGTGACAGCTATCTGGGCGTTTATGAAAACGTTG 469
           ||        | ||  |||  ||| | ||| ||
Db     333 TGTACGGTTGCTGTGTCCGTCGATGCAGACGATG 366


RESULT   9
US-08-472-028A-3/c
; Sequence 3, Application US/08472028A
; Patent No. 5767373
;  GENERAL INFORMATION:
;    APPLICANT:  Ward, Eric R
;    APPLICANT:  Volrath, Sandra
;    TITLE OF INVENTION:  Manipulation of Protoporphyrinogen
;    TITLE OF INVENTION:  Oxidase Enzyme Activity in Eukaryotic Organisms
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/472,028A
;      FILING DATE:
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;      REFERENCE/DOCKET NUMBER:  CGC 1748/CIP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1738 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  70..1596
;      OTHER INFORMATION:  /note= "Arabidopsis protox-2 cDNA;
;      OTHER INFORMATION:  sequence from pWDC-1"
US-08-472-028A-3


  Query Match             2.9%;  Score 30.4;  DB 3;  Length 1738;
  Best Local Similarity   51.5%;  Pred. No. 3.8;
  Matches   70;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;
```

```
;       ZIP:  22202
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/661,610C
;       FILING DATE:  19910228
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Oblon, No. 5292643man F.
;       REGISTRATION NUMBER:  24,618
;       REFERENCE/DOCKET NUMBER:  2292-010-0
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703)521-4500
;       TELEFAX:  (703)486-2347
;       TELEX:  248855 OPAT UR
;   INFORMATION FOR SEQ ID NO:  14:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  624 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  unknown
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  DNA (genomic)
US-07-661-610C-14


  Query Match             3.1%;  Score 31.6;  DB 1;  Length 624;
  Best Local Similarity   49.4%;  Pred. No. 0.79;
  Matches   82;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy     208 GAAGAGAATAATGACGAGCGCACGCTGGCGCGCCAGCAGTTGACCAGCAGCCTGAATGCG 267
           ||  ||   |  | | || |||||||  ||  | | ||   |||  |||||  ||   ||
Db     184 GACCTGATGTACCAGGCGCACACGCTGTCGGCCGACCACCCGACGTGCAGCGCGACGCCG 243

Qy     268 CTGGCGAAGTCCGGCGTGTCATTATCCGCAGAACAAAATGAGAACCTGCGGAGCGCGTTT 327
             || || ||  |||     | |||   |           |||||||   || | |
Db     244 TGCTGGATGTTCGCGGTGCTCGAAACCGGGAATGCGGCCATCGACCTGCGCCACGAGCTG 303

Qy     328 TCTGCGCCGACGTCGGCCTTATTTAGCGCTTCGCCTATGGCGCAGC 373
           |  ||| | |||| | |        |||  || | | | ||  ||
Db     304 GCAACGCTGCCGTCCGACCCGCGCTACGCGCCGACGACGCCGTGGC 349


RESULT   6
US-08-222-616-32
; Sequence 32, Application US/08222616
; Patent No. 5635177
;  GENERAL INFORMATION:
;    APPLICANT:  Bennett, Brian D.
;    APPLICANT:  Goeddel, David
;    APPLICANT:  Lee, James M.
;    APPLICANT:  Matthews, William
;    APPLICANT:  Tsai, Siao Ping
;    APPLICANT:  Wood, William I.
;    TITLE OF INVENTION:  PROTEIN TYROSINE KINASE AGONIST
;    TITLE OF INVENTION:  ANTIBODIES
;    NUMBER OF SEQUENCES:  42
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/222,616
;      FILING DATE:  4-APR-1994
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/00586
;      FILING DATE:  22-JAN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/826935
;      FILING DATE:  22-JAN-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lee, Wendy M.
;      REGISTRATION NUMBER:
;      REFERENCE/DOCKET NUMBER:  821P2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-1994
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  32:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4425 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-222-616-32


  Query Match             3.0%;  Score 30.8;  DB 1;  Length 4425;
  Best Local Similarity   50.0%;  Pred. No. 5.1;
  Matches   77;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy     316 CGGAGCGCGTTTTCTGCGCCGACGTCGGCCTTATTTAGCGCTTCGCCTATGGCGCAGCCG 375
           |||  | ||  | || ||| | ||  |||||   ||    |  | ||| || | |  ||
Db     213 CGGTCCTCGAGTCCTCCGCGGTCGGTGGCCTCTGTTCCTGTCGCTCCTGTGCCCCCACCA 272

Qy     376 AGAACAACCATTTCTGATGCTGAGATTTGGGATATGGTTTCCCAAAATATATCGGCGATA 435
             |  |   |  | | |   ||| | |  |||||   ||| | |||     |   ||   |
Db     273 CGCTCTGACGCTCCCGTGTCTGCGGTCCGGGATGACGTTCCACAACGACGACGTGCTCCA 332

Qy     436 GGTGACAGCTATCTGGGCGTTTATGAAAACGTTG 469
           ||        | ||  |||  ||| | ||| ||
Db     333 TGTACGGTTGCTGTGTCCGTCGATGCAGACGATG 366


RESULT   7
US-08-889-711-1/c
; Sequence 1, Application US/08889711
; Patent No. 5858718
;  GENERAL INFORMATION:
;    APPLICANT:  Holmes, David J.
;    TITLE OF INVENTION:  No. 5858718el pcrA
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dechert Price & Rhoads
;      STREET:  997 Lenox Drive, Building 3, Suite 210
;      CITY:  Lawrenceville
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  08543
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/889,711
;      FILING DATE:
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bloom, Allen
;      REGISTRATION NUMBER:  29,135
;      REFERENCE/DOCKET NUMBER:  P50533
```

```
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01308
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8133 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..8130
US-08-405-496A-5


  Query Match             3.2%;  Score 33.4;  DB 4;  Length 8133;
  Best Local Similarity   49.2%;  Pred. No. 1;
  Matches   88;  Conservative    0;  Mismatches   91;  Indels    0;  Gaps    0;

Qy     824 GAAAAGACTCAAAACTCGAAATGGATAACGCCAAATATCAAGCCTGGCAGTCGGGTTTTA 883
           ||| | |   ||||||  ||   || | |   | | || |  ||   |   |   | |
Db     177 GAATAAATATAAAACTTCAAGCAGAAATAGAGCACTCTCTAATCTAAAAAAAGATATATT 236

Qy     884 AAGCGCAGGAAGAAAATATGAAAACCACATTACAGACGCTGACGCAAAAATATAGCAATG 943
           ||  | ||  ||       | ||||   || ||||  | |||  | |||||  |  || |
Db     237 AAAAGAAGTAATTCTTATTAAAAATTCCAATACAAGCCCTGTAGAAAAAAATTTACATTT 296

Qy     944 CCAATTCATTGTACGACAACCTGGTAAAAGTGCTGAGCAGTACGATAAGTAGCAGCCTG 1002
             ||  || |   || ||    || | | ||||    | |||   ||  |   | |||
Db     297 TGTATGGATAGGTGGAGAAGTCAGTGATATTGCTCTTGAATACATAAAACAATGGGCTG 355


RESULT   4
US-07-661-610C-1
; Sequence 1, Application US/07661610C
; Patent No. 5292643
;  GENERAL INFORMATION:
;    APPLICANT:  Shibano, Yuji
;    APPLICANT:  Toyoda, Hideyoshi
;    APPLICANT:  Utsumi, Ryutaro
;    APPLICANT:  Obata, Kazuaki
;    TITLE OF INVENTION:  Fusaric Acid Resistant Genes
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/661,610C
;      FILING DATE:  19910228
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Oblon, No. 5292643man F.
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  2292-010-0
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5437 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Pseudomonas cepacia
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  519..522
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  531..1832
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  1836..1839
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1847..2275
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  2301..2304
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2314..3354
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  3405..3409
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3414..4040
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  4457..4462
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  4473..4979
US-07-661-610C-1


  Query Match             3.1%;  Score 31.6;  DB 1;  Length 5437;
  Best Local Similarity   49.4%;  Pred. No. 3.2;
  Matches   82;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy     208 GAAGAGAATAATGACGAGCGCACGCTGGCGCGCCAGCAGTTGACCAGCAGCCTGAATGCG 267
           ||   ||   |  | | || ||||||| ||  | | ||   |||  |||||  ||   ||
Db    3597 GACCTGATGTACCAGGCGCACACGCTGTCGGCCGACCACCCGACGTGCAGCGCGACGCCG 3656

Qy     268 CTGGCGAAGTCCGGCGTGTCATTATCCGCAGAACAAAATGAGAACCTGCGGAGCGCGTTT 327
           |  || ||  ||  |||     | |||   |           |||||||   || | |
Db    3657 TGCTGGATGTTCGCGGTGCTCGAAACCGGGAATGCGGCCATCGACCTGCGCCACGAGCTG 3716

Qy     328 TCTGCGCCGACGTCGGCCTTATTTAGCGCTTCGCCTATGGCGCAGC 373
           |  ||| | |||| | |        |||  || | | | ||  ||
Db    3717 GCAACGCTGCCGTCCGACCCGCGCTACGCGCCGACGACGCCGTGGC 3762


RESULT   5
US-07-661-610C-14
; Sequence 14, Application US/07661610C
; Patent No. 5292643
;  GENERAL INFORMATION:
;    APPLICANT:  Shibano, Yuji
;    APPLICANT:  Toyoda, Hideyoshi
;    APPLICANT:  Utsumi, Ryutaro
;    APPLICANT:  Obata, Kazuaki
;    TITLE OF INVENTION:  Fusaric Acid Resistant Genes
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
```

```
Db    1415 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1356

Qy     726 TAGCTGCCTGAAATCTTATGGATCCGGTTATGTCGTCACCGTTGATCTGACGCCATTACA 785
           ::  :   ::::    : :::   ::  : :  :  :  :  ::   :: :  :  : :
Db    1355 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1296

Qy     786 AAAAATGGTTCAGGATATTGATGGTTTAGGCGCGCCGGGAAAAGACTCAAAACTCGAAAT 845
           ::::: ::   :::: :  :: ::   ::: : :  :::::::::   ::::   ::::
Db    1295 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1236

Qy     846 GGATAACGCCAAATATCAAGCCTGGCAGTCGGGTTTTAAAGCGCAGGAAGAAAATATGAA 905
           ::: :: :  ::: :  :::   :: ::  :::    :::: : :::::::::: : :::
Db    1235 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1176

Qy     906 AACCACATTACAGACGCTGACGCAAAAATATAGCAATGCCAATTCATTGTACGACAACCT 965
           ::  : :  : ::: :  :: : ::::: : :: :: :  ::   :  : : :: ::
Db    1175 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1116

Qy     966 GGTAAAAGTGCTGAGCAGTACGATAAGTAGCAGCCTGGAAACCGCCAAAAGC 1017
           :: ::::: :  ::: :: : :: ::: :: ::   :::::  :  :::| |
Db    1115 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRATC 1064


RESULT   2
US-08-480-604A-5
; Sequence 5, Application US/08480604A
; Patent No. 5736139
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,604A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/422,711
;      FILING DATE:  14-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8133 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..8130
US-08-480-604A-5


  Query Match             3.2%;  Score 33.4;  DB 2;  Length 8133;
  Best Local Similarity   49.2%;  Pred. No. 1;
  Matches   88;  Conservative    0;  Mismatches   91;  Indels    0;  Gaps    0;

Qy     824 GAAAAGACTCAAAACTCGAAATGGATAACGCCAAATATCAAGCCTGGCAGTCGGGTTTTA 883
           ||| | |  ||||||  ||   || |  |   | | || |  ||   |   |  | |
Db     177 GAATAAATATAAAACTTCAAGCAGAAATAGAGCACTCTCTAATCTAAAAAAAGATATATT 236

Qy     884 AAGCGCAGGAAGAAAATATGAAAACCACATTACAGACGCTGACGCAAAAATATAGCAATG 943
           ||  | || ||       | ||||   || ||||  | |||  | |||||  |  || |
Db     237 AAAAGAAGTAATTCTTATTAAAAATTCCAATACAAGCCCTGTAGAAAAAAATTTACATTT 296

Qy     944 CCAATTCATTGTACGACAACCTGGTAAAAGTGCTGAGCAGTACGATAAGTAGCAGCCTG 1002
             ||  || |   || ||    || | | ||||    | |||   ||  |   | |||
Db     297 TGTATGGATAGGTGGAGAAGTCAGTGATATTGCTCTTGAATACATAAAACAATGGGCTG 355


RESULT   3
US-08-405-496A-5
; Sequence 5, Application US/08405496A
; Patent No. 5919665
;  GENERAL INFORMATION:
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE FOR CLOSTRIDIUM BOTULINUM
;    TITLE OF INVENTION:  NEUROTOXIN
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/405,496A
;      FILING DATE:  16-MAR-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
```

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 1999, 18:00:51 ; Search time 140.41 Seconds
(without alignments)
764.709 Million cell updates/sec

Title: US-09-068-804-3
Perfect score: 1032
Sequence: 1 ATGCTTAATATTCAAAATTA.....AAAGCTTCCTGCAAGGATAA 1032

Scoring table: IDENTITY_NUC

Searched: 192659 seqs, 52021692 residues

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/5C_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/5D_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	48.8	4.7	7218	2	US-08-232-463-14	Sequence 14, Appl
	2	33.4	3.2	8133	2	US-08-480-604A-5	Sequence 5, Appli
	3	33.4	3.2	8133	4	US-08-405-496A-5	Sequence 5, Appli
	4	31.6	3.1	5437	1	US-07-661-610C-1	Sequence 1, Appli
	5	31.6	3.1	624	1	US-07-661-610C-14	Sequence 14, Appl
	6	30.8	3.0	4425	1	US-08-222-616-32	Sequence 32, Appl
c	7	30.8	3.0	2292	3	US-08-889-711-1	Sequence 1, Appli
	8	30.8	3.0	4425	5	PCT-US95-04228-32	Sequence 32, Appl
c	9	30.4	2.9	1738	3	US-08-472-028A-3	Sequence 3, Appli
c	10	30.4	2.9	1738	4	US-08-808-931-3	Sequence 3, Appli
c	11	30.4	2.9	707	4	US-08-850-910A-40	Sequence 40, Appl
c	12	30.2	2.9	1770	3	US-08-317-305-1	Sequence 1, Appli
c	13	30.2	2.9	1770	5	PCT-US95-12508-1	Sequence 1, Appli
	14	30	2.9	11725	4	US-08-756-506-1	Sequence 1, Appli
c	15	29.8	2.9	3719	3	US-08-920-812-10	Sequence 10, Appl
c	16	29.8	2.9	3719	3	US-08-920-827-10	Sequence 10, Appl
c	17	29.8	2.9	3719	3	US-08-921-177-10	Sequence 10, Appl
c	18	29.8	2.9	3719	3	US-08-362-577C-10	Sequence 10, Appl
	19	29.8	2.9	2319	3	US-08-966-389-5	Sequence 5, Appli
	20	29.8	2.9	3345	3	US-08-966-389-8	Sequence 8, Appli
c	21	29.8	2.9	3719	3	US-08-920-828-10	Sequence 10, Appl
	22	29.8	2.9	2319	4	US-09-103-509-5	Sequence 5, Appli
	23	29.8	2.9	3345	4	US-09-103-509-8	Sequence 8, Appli
	24	29.8	2.9	2319	4	US-09-102-644-5	Sequence 5, Appli
	25	29.8	2.9	3345	4	US-09-102-644-8	Sequence 8, Appli
c	26	29.6	2.9	2017	1	US-07-667-276A-3	Sequence 3, Appli
	27	29.6	2.9	3900	3	US-08-356-060A-42	Sequence 42, Appl
	28	29.2	2.8	2131	1	US-08-325-071-64	Sequence 64, Appl
	29	29.2	2.8	1584	3	US-08-889-711-3	Sequence 3, Appli
	30	29.2	2.8	475	4	US-08-623-906A-4	Sequence 4, Appli
	31	29.2	2.8	1523	4	US-08-713-636-1	Sequence 1, Appli
c	32	29	2.8	2951	3	US-08-386-727-7	Sequence 7, Appli
c	33	28.8	2.8	1478	1	US-08-090-523-27	Sequence 27, Appl
c	34	28.8	2.8	1478	1	US-08-398-627-27	Sequence 27, Appl
c	35	28.8	2.8	1478	1	US-08-406-857-1	Sequence 1, Appli
c	36	28.8	2.8	1478	2	US-08-596-024-3	Sequence 3, Appli
	37	28.8	2.8	2277	2	US-08-676-967-2	Sequence 2, Appli
	38	28.8	2.8	2277	3	US-08-676-974-2	Sequence 2, Appli
	39	28.8	2.8	2277	4	US-09-098-487-2	Sequence 2, Appli
c	40	28.8	2.8	1478	5	PCT-US94-07072-1	Sequence 1, Appli
	41	28.6	2.8	2818	1	US-08-366-276-1	Sequence 1, Appli
c	42	28.6	2.8	476	1	US-08-111-316-2	Sequence 2, Appli
c	43	28.6	2.8	326	1	US-08-111-316-3	Sequence 3, Appli
c	44	28.6	2.8	375	1	US-08-111-316-4	Sequence 4, Appli
c	45	28.6	2.8	326	1	US-08-111-316-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
US-08-232-463-14

Query Match 4.7%; Score 48.8; DB 2; Length 7218;
Best Local Similarity 5.6%; Pred. No. 8.4e-06;
Matches 23; Conservative 216; Mismatches 173; Indels 0; Gaps 0;

```
Qy    606 CAATAAATATAATCAAATAAACAGTAATACCGTTTTATTTCCAGCGCAGTCAGGCAGCGG 665
          | ||  ||  ||  |  |||| ||  | |    |||  | | :: : ::  ::: :: ::
Db   1475 CTATCTATGCAAGTAGTTAAAGAGATAGAAGAATTTGGTACRRRRRRRRRRRRRRRRRRR 1416

Qy    666 CGTTAAAGTAGCCACTGAAGCGGAAGCGAGACAGTGGCTCAGTGAATTGAATTTACCGAA 725
            :  :::: ::  :  :::: ::::: :::: :: ::   :: :::  :::   :  :::
```

/clone_lib="Soares ovary tumor NbHOT"
/sex="Female"
/tissue_type="ovarian tumor"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: ovary; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCGGTTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). Library constructed by Bento Soares and M.Fatima Bonaldo. "

BASE COUNT 81 a 74 c 90 g 58 t
ORIGIN

Query Match 3.4%; Score 35.2; DB 33; Length 303;
Best Local Similarity 53.7%; Pred. No. 6.8;
Matches 73; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

```
Qy    581 TCAAAAATGATTTAAACAGTTTAGTCAATAAATATAATCAAATAAACAGTAATACCGTTT 640
          | |||||| |  | ||||  | | || || ||| | ||  ||| |||  | || ||  | |
Db      8 TTAAAAAAGCCIGAAACTTTATTGTAAAAAAAAAAAAAAAAAGAAAATGAAAAACACTCT 67

Qy    641 TATTTCCAGCGCAGTCAGGCAGCGGCGTTAAAGTAGCCACTGAAGCGGAAGCGAGACAGT 700
            |   || | || | | ||  | | ||| | ||     |  ||   |||   |   | ||
Db     68 AAAACCCTGAGCTGGCTGGGGGTGTCGTCAGAGATCAGAAGGACAGGGAGAAGGCTCGGT 127

Qy    701 GGCTCAGTGAATTGAA 716
            | |||    ||||
Db    128 AACCCAGGTTGGTGAA 143
```

RESULT 15
AI283105/c
LOCUS AI283105 429 bp mRNA EST 29-JAN-1999
DEFINITION qm51g04.x1 Soares_placenta_8to9weeks_2NbHP8to9W Homo sapiens cDNA clone IMAGE:1892310 3', mRNA sequence.
ACCESSION AI283105
NID g3921338
VERSION AI283105.1 GI:3921338
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 429)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Nov 29, 1993 this sequence version replaced gi:636945.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 636 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 375.

FEATURES Location/Qualifiers
source 1. .429
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1892310"
/clone_lib="Soares_placenta_8to9weeks_2NbHP8to9W"
/dev_stage="two placentae: one from 8 weeks and another from 9 weeks post conception"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: placenta; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCGATTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). Library constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 111 a 70 c 80 g 168 t
ORIGIN

Query Match 3.4%; Score 35.2; DB 44; Length 429;
Best Local Similarity 58.7%; Pred. No. 7.4;
Matches 61; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

```
Qy    918 GACGCTGACGCAAAAATATAGCAATGCCAATTCATTGTACGACAACCTGGTAAAAGTGCT 977
          | || | |   | ||  |||  ||    |||| ||||   ||   |||| |||||| |||
Db    313 GCCGGTTATAAACAAGAATAAAAACTTGAATTTATTGGTGGAAGTCCTGTTAAAAGAGCT 254

Qy    978 GAGCAGTACGATAAGTAGCAGCCTGGAAACCGCCAAAAGCTTCC 1021
          | |   |  ||| |   ||  ||| |||  || || || ||||
Db    253 CAACCCAAGCATATGAGACAAACTGCAAAAAGCAAACAGTTTCC 210
```

Search completed: October 5, 1999, 17:13:57
Job time: 8680 sec

```
                     /db_xref="taxon:10116"
                     /clone="UI-R-A0-ar-b-11-0-UI"
                     /clone_lib="UI-R-A0"
                     /dev_stage="adult"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; This library
                     consists of a mixture of individually tagged normalized
                     libraries constructed from rat placenta, adult lung,
                     brain, liver, kidney, heart, spleen, ovary, and muscle.
                     The tag is a string of 3-5 nucleotides present between the
                     Not I site and the oligo-dT track which allows
                     identification of the library of origin of a clone within
                     the mixture."
BASE COUNT      146 a    115 c    102 g    175 t
ORIGIN


  Query Match             3.5%;  Score 36.2;  DB 45;  Length 538;
  Best Local Similarity   48.8%;  Pred. No. 4;
  Matches   98;  Conservative    0;  Mismatches  103;  Indels    0;  Gaps    0;

Qy     527 GCTGGTTATTACCAGGTAAGGACGGTAATACCGTTAAGCTAGATGTTACCTCACTCAAAA 586
           || || || ||       ||| | ||  |||||| ||| || ||| | | ||
Db     209 GCAGGCTAGTATATACCGAGGGGGATACAACCGTTTTGCTCCATATTAAATGATAAAACC 150

Qy     587 ATGATTTAAACAGTTTAGTCAATAAATATAATCAAATAAACAGTAATACCGTTTTATTTC 646
           || |   |||||     |   | ||| | || |||| ||||     | ||  | |
Db     149 ATTAAACAAACAAACAAAAAACAAAACAAAAACAAAAAAACCAACCTTCCAATGTGGGGA 90

Qy     647 CAGCGCAGTCAGGCAGCGGCGTTAAAGTAGCCACTGAAGCGGAAGCGAGACAGTGGCTCA 706
           || | |  |   | | |||   |  || || ||  | || | ||     || |    ||
Db      89 GAGAGGAAGCTTTCCGAGGCCCGAGTGTTGCGACACATGCAGTAGGACATCACTTTAGCA 30

Qy     707 GTGAATTGAATTTACCGAATA 727
               |  |||   ||  || |
Db      29 ACTCAAAGAAACAACTAAAAA 9


RESULT  13
AA368794
LOCUS       AA368794      382 bp    mRNA            EST       21-APR-1997
DEFINITION  EST80100 Placenta I Homo sapiens cDNA similar to plasma protein-A,
            pregnancy associated, mRNA sequence.
ACCESSION   AA368794
NID         g2021112
VERSION     AA368794.1  GI:2021112
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 382)
  AUTHORS   Adams,M.D., Soares,M.B., Kerlavage,A.R., Fields,C. and Venter,J.C.
  TITLE     Rapid cDNA sequencing (expressed sequence tags) from a
            directionally cloned human infant brain cDNA library
  JOURNAL   Nature Genet. 4, 373-380 (1993)
  MEDLINE   94004965
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1393000.
            Other_ESTs: THC169181
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .382
                     /organism="Homo sapiens"
                     /db_xref="ATCC (inhost):173490"
                     /db_xref="taxon:9606"
                     /map="788H12; 14q24.3"
                     /clone_lib="Placenta I"
                     /tissue_type="placenta"
                     /dev_stage="fetus"
                     /note="Organ: placenta; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: EcoRI"
BASE COUNT      90 a    103 c    103 g     75 t     11 others
ORIGIN


  Query Match             3.5%;  Score 35.8;  DB 32;  Length 382;
  Best Local Similarity   50.3%;  Pred. No. 4.9;
  Matches   82;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy     127 ACAGATATCATTTCATTATCGCAGGCGGCTACTAAAATCCACCAGGCACAGCAGACGCTG 186
           |||||  ||||| |    |   |  |   | |   ||| |||| |  ||  |  ||
Db      73 ACAGAGGTCATTGCCAGNTACAATCAGCTCTCAAGTTTCCGCCAGCCCAAGGTGGTGCGN 132

Qy     187 CAGTCAACGCCACCGATCTCTGAAGAGAATAATGACGAGCGCACGCTGGCGCGCCAGCAG 246
           |     |       ||| |||||| || || |  | |  ||| || |||||  | |||
Db     133 TACCGCGTGGTCAACNTCTATGAAGATGATCATAAGAACCCGACGGTGACGCGCGANCAG 192

Qy     247 TTGACCAGCAGCCTGAATGCGCTGGCGAAGTCCGGCGTGTCAT 289
           || |  |   |   || ||||||  || ||  |  |  ||
Db     193 GTGGACTTCCAGCACCATCAGCTGGCTGAGGCCTTCAAGCAAT 235


RESULT  14
AA404589
LOCUS       AA404589      303 bp    mRNA            EST       08-AUG-1997
DEFINITION  zt43a03.s1 Soares ovary tumor NbHOT Homo sapiens cDNA clone
            IMAGE:725068 3', mRNA sequence.
ACCESSION   AA404589
NID         g2058801
VERSION     AA404589.1  GI:2058801
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 303)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,
            Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
            Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
            Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J.,
            Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.
            and Marra,M.
  TITLE     Generation and analysis of 280,000 human expressed sequence tags
  JOURNAL   Genome Res. 6 (9), 807-828 (1996)
  MEDLINE   97044478
COMMENT     On May 8, 1995 this sequence version replaced gi:801092.

            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 388   Std Error: 0.00
            Seq primer: -41m13 fwd. ET from Amersham.
FEATURES             Location/Qualifiers
     source          1. .303
                     /organism="Homo sapiens"
                     /db_xref="GDB:5936815"
                     /db_xref="taxon:9606"
                     /map="3"
                     /clone="IMAGE:725068"
```

IGTTACCAATCTGAAGTGGGAGCGGCCGCATAGGTTTTTTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
Library is normalized, and was constructed by Bento
Soares and M.Fatima Bonaldo."
BASE COUNT 94 a 81 c 100 g 88 t
ORIGIN

Query Match 3.6%; Score 37; DB 42; Length 363;
Best Local Similarity 54.9%; Pred. No. 2.2;
Matches 73; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

Qy 584 AAAATGATTTAAACAGTTTAGTCAATAAATATAATCAAATAAACAGTAATACCGTTTTAT 643
Db 28 AAAAAGCCTGAAACTTTATTGTAAAAAAAAAAAAAAAAAAGAAAATGAAAAACACTTTAAA 87

Qy 644 TTCCAGCGCAGTCAGGCAGCGGCGTTAAAGTAGCCACTGAAGCGGAAGCGAGACAGTGGC 703
Db 88 ACCCTGAGCTGGCTGGGGGTGTCGTCAGAGATCAGAAGGACAGGGAAAAGGCTCGGTAAC 147

Qy 704 TCAGTGAATTGAA 716
Db 148 CCAGGTTGGTGAA 160

RESULT 11
AI021088/c
LOCUS AI021088 337 bp mRNA EST 16-JUN-1998
DEFINITION ub01b01.r1 Soares mouse mammary gland NbMMG Mus musculus cDNA clone IMAGE:1365673 5', mRNA sequence.
ACCESSION AI021088
NID g3235424
VERSION AI021088.1 GI:3235424
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 337)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Jan 19, 1998 this sequence version replaced gi:2151460.

Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:898893
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 330.
FEATURES Location/Qualifiers
source 1. .337
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:1365673"
/clone_lib="Soares mouse mammary gland NbMMG"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATGGTTTTTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. RNA provided by Dr. Minoru Ko, Wayne State Univ. Library constructed and normalized by Bento Soares and M.Fatima Bonaldo."
BASE COUNT 102 a 53 c 61 g 121 t
ORIGIN

Query Match 3.6%; Score 36.8; DB 41; Length 337;
Best Local Similarity 55.5%; Pred. No. 2.4;
Matches 71; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

Qy 476 TCTATACCGATTTTTATCAGGCCTTCAGTGATATTCTTTCCAAAATGGGAGGCTGGTTAT 535
Db 170 TCTGAAAATACATTTAAAAAAAACTCATTTAGTGTTAGGGTAAAATAGAACCCTTATTTG 111

Qy 536 TACCAGGTAAGGACGGTAATACCGTTAAGCTAGATGTTACCTCACTCAAAAATGATTTAA 595
Db 110 TAACCATTCATTACAGACAAACCAATAAGGAAGAAACTAGGTCACCTAAAAAAGTTTTAA 51

Qy 596 ACAGTTTA 603
Db 50 ACAGTTTA 43

RESULT 12
AA818804/c
LOCUS AA818804 538 bp mRNA EST 05-FEB-1999
DEFINITION UI-R-A0-ar-b-11-0-UI.s1 UI-R-A0 Rattus norvegicus cDNA clone UI-R-A0-ar-b-11-0-UI 3' similar to gi|2825286|gb|AA773715|AA773715 af81c05.r1 Soares NhHMPu S1 Homo sapiens cDNA clone 1048424 5', mRNA sequence.
ACCESSION AA818804
NID g4227480
VERSION AA818804.1 GI:4227480
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 538)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT On Feb 17, 1998 this sequence version replaced gi:2888390.

Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized adult Brain library. cDNA Library Preparation: M. Fatima Bonaldo, Ph.D. Clone distribution: clones will be available through Research Genetics The following repetitive elements were found in this cDNA sequence: 105-154, >(CAAA)n#Simple_repeat
Seq primer: M13 Forward.
FEATURES Location/Qualifiers
source 1. .538
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"

```
  Query Match             3.7%;  Score 38;  DB 39;  Length 349;
  Best Local Similarity   52.5%;  Pred. No. 1.1;
  Matches   83;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy     488 TTTATCAGGCCTTCAGTGATATTCTTTCCAAAATGGGAGGCTGGTTATTACCAGGTAAGG 547
           ||||  |     ||| | |   |       ||||| | |  ||  ||  || |   | |
Db     345 TTTAAAAAAAACTCATTTAGTGTTAGGGTAAAATAGAACCCTTATTTGTAACCATTCATT 286

Qy     548 ACGGTAATACCGTTAAGCTAGATGTTACCTCACTCAAAAATGATTTAAACAGTTTAGTCA 607
           || |  | |||  ||||  |||   ||  ||||  ||||| | ||||||||||||||
Db     285 ACAGACAAACCAATAAGGAAGAAACTAGGTCACCTAAAAAAGTTTTAAACAGTTTATATG 226

Qy     608 ATAAATATAATCAAATAAACAGTAATACCGTTTTATTT 645
           |   | | |     |  | |   | ||  |||| |||
Db     225 AATGAGAACACATGCTTTAAACACAAACTATTTTTTTT 188


RESULT  9
AA260634/c
LOCUS       AA260634      292 bp    mRNA            EST       18-MAR-1997
DEFINITION  va11f02.r1 Soares mouse lymph node NbMLN Mus musculus cDNA clone
            IMAGE:722619 5', mRNA sequence.
ACCESSION   AA260634
NID         g1897136
VERSION     AA260634.1  GI:1897136
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 292)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1402250.

            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:448115
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 272.
FEATURES             Location/Qualifiers
     source          1. .292
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:722619"
                     /clone_lib="Soares mouse lymph node NbMLN"
                     /sex="male"
                     /tissue_type="lymph node"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
                     /note="Organ: lymph node; Vector: pT7T3D-Pac (Pharmacia)
                     with a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     1st strand cDNA was primed with a Not I - oligo(dT) primer
                     [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGATACTTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. RNA
                     provided by Dr. Bertrand Jordan. Library constructed and
                     normalized by Bento Soares and M.Fatima Bonaldo."
BASE COUNT       82 a     43 c     56 g    111 t
ORIGIN


  Query Match             3.6%;  Score 37.2;  DB 30;  Length 292;
  Best Local Similarity   51.2%;  Pred. No. 1.8;
  Matches   87;  Conservative    0;  Mismatches   83;  Indels    0;  Gaps    0;

Qy     476 TCTATACCGATTTTTATCAGGCCTTCAGTGATATTCTTTCCAAAATGGGAGGCTGGTTAT 535
           |||  |    |  ||||  |      ||| | |   |       |||||  | |  ||  ||
Db     179 TCTGAAAATACATTTAAAAAAAACTCATTTAGTGTTAGGGTAAAATAGAACCCTTATTTG 120

Qy     536 TACCAGGTAAGGACGGTAATACCGTTAAGCTAGATGTTACCTCACTCAAAAATGATTTAA 595
           || |   | |  || |  | |||  | ||  |||   ||  ||||  |||||  | |||||
Db     119 TAACCATTCATTACAGACAAACCAATTAGGAAGAAACTAGGTCACCTAAAAAAGTTTTAA 60

Qy     596 ACAGTTTAGTCAATAAATATAATCAAATAAACAGTAATACCGTTTTATTT 645
           ||||||||    |   | |  |     |  | |   | ||  |||| |||
Db      59 ACAGTTTATATGAATGAGAACACATGCTTTAAACACAAACTATTTTTTTT 10


RESULT  10
AI083764
LOCUS       AI083764      363 bp    mRNA            EST       13-FEB-1999
DEFINITION  qf23c06.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:1750858 3'
            similar to contains element MER22 repetitive element ;, mRNA
            sequence.
ACCESSION   AI083764
NID         g3422187
VERSION     AI083764.1  GI:3422187
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 363)
  AUTHORS   NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute / National Institute of Neurological
            Disorders and Stroke, Brain Tumor Genome Anatomy Project
            (CGAP/BTGAP), Tumor Gene Index
  JOURNAL   Unpublished (1998)
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2152392.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
            Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 548   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 255.
FEATURES             Location/Qualifiers
     source          1. .363
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="19p13.1"
                     /clone="IMAGE:1750858"
                     /clone_lib="NCI_CGAP_Brn25"
                     /tissue_type="anaplastic oligodendroglioma"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
```

```
  Query Match             3.8%;  Score 38.8;  DB 35;  Length 555;
  Best Local Similarity   51.8%;  Pred. No. 0.73;
  Matches   88;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy     476 TCTATACCGATTTTTATCAGGCCTTCAGTGATATTCTTTCCAAAATGGGAGGCTGGTTAT 535
           |||  |   |  ||||  |     ||| | |   |      |||||  | |  ||  ||
Db     209 TCTGAAAATACATTTAAAAAAAACTCATTTAGTGTTAGGGTAAAATAGAACCCTTATTTG 150

Qy     536 TACCAGGTAAGGACGGTAATACCGTTAAGCTAGATGTTACCTCACTCAAAAATGATTTAA 595
           ||  |   | |  ||  |  |||  ||||  |||   ||  ||||  ||||| | |||||
Db     149 TAACCATTCATTACAGACAAACCAATAAGGAAGAAACTAGGTCACCTAAAAAAGTTTTAA 90

Qy     596 ACAGTTTAGTCAATAAATATAATCAAATAAACAGTAATACCGTTTTATTT 645
           ||||||||    |   | |  |    |  | |   | ||  ||||  |||
Db      89 ACAGTTTATATGAATGAGAACACATGCTTTAAACACAAACTATTTTTTTT 40


RESULT   7
AI197155/c
LOCUS       AI197155      369 bp    mRNA            EST       14-OCT-1998
DEFINITION  ue51h03.r1 Soares mouse mammary gland NMLMG Mus musculus cDNA clone
            IMAGE:1494677 5', mRNA sequence.
ACCESSION   AI197155
NID         g3749761
VERSION     AI197155.1  GI:3749761
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 369)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Aug 21, 1998 this sequence version replaced.

            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:932281
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 325.
FEATURES             Location/Qualifiers
     source          1. .369
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1494677"
                     /clone_lib="Soares mouse mammary gland NMLMG"
                     /sex="female (lactating)"
                     /tissue_type="mammary gland"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; 1st strand cDNA was prepared from mammary
                     gland tissue from a lactating female, and was then primed
                     with a Not I - oligo(dT) primer. Double-stranded  cDNA was
                     ligated to Eco RI adaptors (Pharmacia), digested with Not
                     I  and cloned into the Not I and Eco RI sites of the
                     modified pT7T3 vector.  Library is normalized.  Library
                     was constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      108 a     57 c     70 g    134 t
ORIGIN
```

```
  Query Match             3.8%;  Score 38.8;  DB 43;  Length 369;
  Best Local Similarity   51.8%;  Pred. No. 0.66;
  Matches   88;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy     476 TCTATACCGATTTTTATCAGGCCTTCAGTGATATTCTTTCCAAAATGGGAGGCTGGTTAT 535
           |||  |   |  ||||  |     ||| | |   |      |||||  | |  ||  ||
Db     202 TCTGAAAATACATTTAAAAAAAACTCATTTAGTGTTAGGGTAAAATAGAACCCTTATTTG 143

Qy     536 TACCAGGTAAGGACGGTAATACCGTTAAGCTAGATGTTACCTCACTCAAAAATGATTTAA 595
           ||  |   | |  ||  |  |||  ||||  |||   ||  ||||  ||||| | |||||
Db     142 TAACCATTCATTACAGACAAACCAATAAGGAAGAAACTAGGTCACCTAAAAAAGTTTTAA 83

Qy     596 ACAGTTTAGTCAATAAATATAATCAAATAAACAGTAATACCGTTTTATTT 645
           ||||||||    |   | |  |    |  | |   | ||  ||||  |||
Db      82 ACAGTTTATATGAATGAGAACACATGCTTTAAACACAAACTATTTTTTTT 33


RESULT   8
AA839883/c
LOCUS       AA839883      349 bp    mRNA            EST       27-FEB-1998
DEFINITION  vw65d05.r1 Stratagene mouse heart (#937316) Mus musculus cDNA clone
            IMAGE:1259817 5', mRNA sequence.
ACCESSION   AA839883
NID         g2916477
VERSION     AA839883.1  GI:2916477
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 349)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2153144.

            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:662369
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 346.
FEATURES             Location/Qualifiers
     source          1. .349
                     /organism="Mus musculus"
                     /strain="NIH/Swiss"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1259817"
                     /clone_lib="Stratagene mouse heart (#937316)"
                     /sex="pooled"
                     /tissue_type="heart"
                     /dev_stage="13 day embryos"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Organ: heart; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. 93 pooled NIH/Swiss 13 day embryo hearts.
                     Average insert size: 1.0 kb; Uni-ZAP XR Vector; ~5'
                     adaptor sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor
                     sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT       90 a     53 c     60 g    146 t
ORIGIN
```

```
Qy     476 TCTATACCGATTTTTATCAGGCCTTCAGTGATATTCTTTCCAAAATGGGAGGCTGGTTAT 535
           |||  |   |  ||||  |      ||| | |   |       ||||| | |  ||  ||
Db     290 TCTGAAAATACATTTAAAAAAAACTCATTTAGTGTTAGGGTAAAATAGAACCCTTATTTG 231

Qy     536 TACCAGGTAAGGACGGTAATACCGTTAAGCTAGATGTTACCTCACTCAAAAATGATTTAA 595
           || |   | |  || |  | |||  ||||  |||   ||  ||||  ||||| | |||||
Db     230 TAACCATTCATTACAGACAAACCAATAAGGAAGAAACTAGGTCACCTAAAAAAGTTTTAA 171

Qy     596 ACAGTTTAGTCAATAAATATAATCAAATAAACAGTAATACCGTTTTATTT 645
           ||||||||    |  | |  |     | | |   | ||  |||| |||
Db     170 ACAGTTTATATGAATGAGAACACATGCTTTAAACACAAACTATTTTTTTT 121


RESULT  5
AA475503/c
LOCUS       AA475503      445 bp    mRNA            EST       18-JUN-1997
DEFINITION  vh15f12.r1 Soares mouse mammary gland NbMMG Mus musculus cDNA clone
            IMAGE:875567 5', mRNA sequence.
ACCESSION   AA475503
NID         g2203354
VERSION     AA475503.1  GI:2203354
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 445)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On May 9, 1995 this sequence version replaced gi:802532.

            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:515047
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 401.
FEATURES             Location/Qualifiers
     source          1. .445
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:875567"
                     /clone_lib="Soares mouse mammary gland NbMMG"
                     /sex="male"
                     /tissue_type="mammary gland"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
                     /note="Organ: mammary gland; Vector: pT7T3D-Pac
                     (Pharmacia) with a modified polylinker; Site_1: Not I;
                     Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                     oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATGGTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
                     constructed and normalized by Bento Soares and M.Fatima
                     Bonaldo."
BASE COUNT      123 a     72 c     81 g    169 t
ORIGIN
```

```
  Query Match             3.8%;  Score 38.8;  DB 34;  Length 445;
  Best Local Similarity   51.8%;  Pred. No. 0.69;
  Matches   88;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy     476 TCTATACCGATTTTTATCAGGCCTTCAGTGATATTCTTTCCAAAATGGGAGGCTGGTTAT 535
           |||  |   |  ||||  |      ||| | |   |       ||||| | |  ||  ||
Db     281 TCTGAAAATACATTTAAAAAAAACTCATTTAGTGTTAGGGTAAAATAGAACCCTTATTTG 222

Qy     536 TACCAGGTAAGGACGGTAATACCGTTAAGCTAGATGTTACCTCACTCAAAAATGATTTAA 595
           || |   | |  || |  | |||  ||||  |||   ||  ||||  ||||| | |||||
Db     221 TAACCATTCATTACAGACAAACCAATAAGGAAGAAACTAGGTCACCTAAAAAAGTTTTAA 162

Qy     596 ACAGTTTAGTCAATAAATATAATCAAATAAACAGTAATACCGTTTTATTT 645
           ||||||||    |  | |  |     | | |   | ||  |||| |||
Db     161 ACAGTTTATATGAATGAGAACACATGCTTTAAACACAAACTATTTTTTTT 112


RESULT  6
AA545466/c
LOCUS       AA545466      555 bp    mRNA            EST       04-AUG-1997
DEFINITION  vj99g12.r1 Knowles Solter mouse blastocyst B1 Mus musculus cDNA
            clone IMAGE:945286 5', mRNA sequence.
ACCESSION   AA545466
NID         g2306540
VERSION     AA545466.1  GI:2306540
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 555)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On May 5, 1995 this sequence version replaced gi:798612.

            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:542142
            High quality sequence stop: 352.
FEATURES             Location/Qualifiers
     source          1. .555
                     /organism="Mus musculus"
                     /strain="B6D2 F1/J"
                     /db_xref="taxon:10090"
                     /map="21"
                     /clone="IMAGE:945286"
                     /clone_lib="Knowles Solter mouse blastocyst B1"
                     /tissue_type="blastocyst"
                     /dev_stage="embryo (pre-implantation)"
                     /lab_host="DH10B"
                     /note="Organ: embryo; Vector: pSPORT; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally from mRNA prepared
                     from 800 blastocysts.  Primer: SalI(dT):
                     5'-CGGTCGACCGTCGACCGTTTTTTTTTTTTTT-3'.  cDNAs were
                     cloned into the NotI/SalI sites of a pSPORT vector (Life
                     Technologies).  Two different size selections:  B1 (larger
                      inserts) and B3."
BASE COUNT      156 a     100 c     114 g    185 t
ORIGIN
```

```
RESULT   3
W57781
LOCUS       W57781         325 bp    mRNA            EST       15-OCT-1996
DEFINITION  zd17b03.r1 Soares_fetal_heart_NbHH19W Homo sapiens cDNA clone
            IMAGE:340877 5' similar to gb:X52104 P68 PROTEIN (HUMAN);, mRNA
            sequence.
ACCESSION   W57781
NID         g1364545
VERSION     W57781.1  GI:1364545
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 325)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     On Apr 5, 1995 this sequence version replaced gi:760479.

            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1879   Std Error: 0.00
            Seq primer: mob.REGA+ET
            High quality sequence stop: 247.
FEATURES             Location/Qualifiers
     source          1. .325
                     /organism="Homo sapiens"
                     /db_xref="GDB:1266252"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:340877"
                     /clone_lib="Soares_fetal_heart_NbHH19W"
                     /sex="unknown"
                     /dev_stage="19 weeks"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: heart; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATCTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by
                     M.Fatima Bonaldo. This library was constructed from the
                     same fetus as the fetal lung library, Soares fetal lung
                     NbHL19W."
BASE COUNT      113 a     50 c     75 g     83 t      4 others
ORIGIN


  Query Match             3.8%;  Score 38.8;  DB 26;  Length 325;
  Best Local Similarity   49.7%;  Pred. No. 0.64;
  Matches   97;  Conservative    0;  Mismatches   98;  Indels    0;  Gaps    0;

Qy     718 TTACCGAATAGCTGCCTGAAATCTTATGGATCCGGTTATGTCGTCACCGTTGATCTGACG 777
           || | ||     | |||||||   |||          |  | |   |  |||| ||||
Db     120 TTGCTGAGAGATTTCCTGAAAGACTATATTCATATAAACATTGGTGCACTTGAACTGAGT 179

Qy     778 CCATTACAAAAAATGGTTCAGGATATTGATGGTTTAGGCGCGCCGGGAAAAGACTCAAAA 837
           ||   || || ||  |||||  | |||   || |       |   ||||||  | ||||
Db     180 GCAAACCACAACATTCTTCAGATTGTTGGATGTGTGTCATGACGTAGAAAAGGATGAAAA 239

Qy     838 CTCGAAATGGATAACGCCAAATATCAAGCCTGGCAGTCGGGTTTTAAAGCGCAGGAAGAA 897
           |   |  ||| |  | ||||||  |  ||  |  || | || ||| |    |
Db     240 ACTTATTCGTCTAATGGAAGATATCATGAGTGAGAAACGAGATTAAAACCATTTGNTTTT 299

Qy     898 AATATGAAAACCACA 912
           | || |||| | | |
Db     300 AGTAAGAAACCAAAA 314


RESULT   4
AA174511/c
LOCUS       AA174511       513 bp    mRNA            EST       16-FEB-1997
DEFINITION  ms69h04.r1 Soares mouse 3NbMS Mus musculus cDNA clone IMAGE:616855
            5', mRNA sequence.
ACCESSION   AA174511
NID         g1755634
VERSION     AA174511.1  GI:1755634
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 513)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Apr 14, 1993 this sequence version replaced gi:692920.

            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:377679
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 493.
FEATURES             Location/Qualifiers
     source          1. .513
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /map="1"
                     /clone="IMAGE:616855"
                     /clone_lib="Soares mouse 3NbMS"
                     /sex="male"
                     /tissue_type="Spleen"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGCTGTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. RNA
                     provided by Dr. Bertrand Jordan. Library went through
                     three rounds of normalization, and was constructed by
                     Bento Soares and M.Fatima Bonaldo."
BASE COUNT      139 a     85 c     100 g     189 t
ORIGIN


  Query Match             3.8%;  Score 38.8;  DB 29;  Length 513;
  Best Local Similarity   51.8%;  Pred. No. 0.72;
  Matches   88;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;
```

```
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 323)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1393272.

            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:407699
            Putative full length read
            vector to vector length is 515
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 306.
FEATURES             Location/Qualifiers
     source          1. .323
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:661851"
                     /clone_lib="GuayWoodford Beier mouse kidney day 7"
                     /tissue_type="kidney"
                     /dev_stage="juvenile (7 days old)"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Organ: kidney; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. Average insert size: 1.0 kb; Uni-ZAP XR Vector;
                     ~5' adaptor sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor
                     sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3' Library provided
                     by Lisa Guay-Woodford."
BASE COUNT      128 a     58 c     52 g     85 t
ORIGIN
```

```
  Query Match             3.9%;  Score 40;  DB 30;  Length 323;
  Best Local Similarity   57.0%;  Pred. No. 0.29;
  Matches   73;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy     476 TCTATACCGATTTTTATCAGGCCTTCAGTGATATTCTTTCCAAAATGGGAGGCTGGTTAT 535
           |||  |   |  ||||  |     ||| | | | |     |||||| | |  ||  |||
Db     159 TCTGAAAATACATTTAAAAAAAACTCATTTAGAGTTAGGGAAAAATAGAACCCTTATTAG 218

Qy     536 TACCAGGTAAGGACGGTAATACCGTTAAGCTAGATGTTACCTCACTCAAAAATGATTTAA 595
           || |   | |  || |  | |||  ||||  |||   ||  ||||  ||||| | |||||
Db     219 TAACCAATCATTACAGACAAACCAATAAGGAAGAAACTAGGTCACCTAAAAAAGTTTTAA 278

Qy     596 ACAGTTTA 603
           ||||||||
Db     279 ACAGTTTA 286
```

```
RESULT   2
W54408/c
LOCUS       W54408         423 bp    mRNA            EST       03-JUN-1996
DEFINITION  md02h05.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
            clone IMAGE:367257 5', mRNA sequence.
ACCESSION   W54408
NID         gi1355450
VERSION     W54408.1  GI:1355450
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 423)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Apr 14, 1993 this sequence version replaced gi:692592.

            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:228689
            Seq primer: ETPrimer
            High quality sequence stop: 345.
FEATURES             Location/Qualifiers
     source          1. .423
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /map="7"
                     /clone="IMAGE:367257"
                     /clone_lib="Soares mouse embryo NbME13.5 14.5"
                     /sex="unknown"
                     /tissue_type="embryo"
                     /dev_stage="13.5-14.5dpc total fetus"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGGAAATTTTTTTTTTTTTTTTTTTTTT
                     T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
                     14.5dpc embryos [total RNA provided by  Minoru Ko, Wayne
                     State Univ., from 2 ]; double-stranded cDNA was ligated to
                     Eco RI adaptors (Pharmacia), digested with Not I and
                     cloned into the Not I and Eco RI sites of the modified
                     pT7T3 vector. Library went through one round of
                     normalization, and was constructed by Bento Soares and
                     M.Fatima Bonaldo.  "
BASE COUNT      121 a     66 c     74 g     162 t
ORIGIN
```

```
  Query Match             3.8%;  Score 38.8;  DB 26;  Length 423;
  Best Local Similarity   51.8%;  Pred. No. 0.68;
  Matches   88;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy     476 TCTATACCGATTTTTATCAGGCCTTCAGTGATATTCTTTCCAAAATGGGAGGCTGGTTAT 535
           |||  |   |  ||||  |     ||| | |   |     ||||| | |  ||  ||
Db     257 TCTGAAAATACATTTAAAAAAAACTCATTTAGTGTTAGGGTAAAATAGAACCCTTATTTG 198

Qy     536 TACCAGGTAAGGACGGTAATACCGTTAAGCTAGATGTTACCTCACTCAAAAATGATTTAA 595
           || |   | |  || |  | |||  ||||  |||   ||  ||||  ||||| | |||||
Db     197 TAACCATTCATTACAGACAAACCAATAAGGAAGAAACTAGGTCACCTAAAAAAGTTTTAA 138

Qy     596 ACAGTTTAGTCAATAAATATAATCAAATAAACAGTAATACCGTTTIATTT 645
           ||||||||    |   | |  |    | | |  | ||  |||| |||
Db     137 ACAGTTTATATGAATGAGAACACATGCTTTAAACACAAACTATTTTTTTT 88
```

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 1999, 17:13:51 ; Search time 1495.71 Seconds
(without alignments)
1360.995 Million cell updates/sec

Title: US-09-068-804-3
Perfect score: 1032
Sequence: 1 ATGCTTAATATTCAAAATTA.....AAAGCTTCCTGCAAGGATAA 1032

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 986266752 residues

Database : EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		40	3.9	323	30	AA208468	AA208468 mv85f02.r
c 2		38.8	3.8	423	26	W54408	W54408 md02h05.r1
3		38.8	3.8	325	26	W57781	W57781 zd17b03.r1
c 4		38.8	3.8	513	29	AA174511	AA174511 ms69h04.r
c 5		38.8	3.8	445	34	AA475503	AA475503 vh15f12.r
c 6		38.8	3.8	555	35	AA545466	AA545466 vj99g12.r
c 7		38.8	3.8	369	43	AI197155	AI197155 ue51h03.r
c 8		38	3.7	349	39	AA839883	AA839883 vw65d05.r
c 9		37.2	3.6	292	30	AA260634	AA260634 va11f02.r
10		37	3.6	363	42	AI083764	AI083764 qf23c06.x
c 11		36.8	3.6	337	41	AI021088	AI021088 ub01b01.r
c 12		36.2	3.5	538	45	AA818804	AA818804 UI-R-A0-a
13		35.8	3.5	382	32	AA368794	AA368794 EST80100
14		35.2	3.4	303	33	AA404589	AA404589 zt43a03.s
c 15		35.2	3.4	429	44	AI283105	AI283105 qm51g04.x
16		34.8	3.4	648	29	AA126510	AA126510 zn86b12.s
17		34.4	3.3	302	20	T09697	T09697 0287m3 gmbP
18		34.4	3.3	400	25	N56849	N56849 yw83f02.r1
19		34	3.3	397	44	AI267656	AI267656 aq92g10.x
c 20		34	3.3	496	46	AA924501	AA924501 UI-R-A1-e
21		33.8	3.3	405	21	R12077	R12077 yf54g12.r1
22		33.8	3.3	241	28	AA092922	AA092922 m0602.seq
23		33.8	3.3	333	39	AA825589	AA825589 od57g01.s
24		33.4	3.2	391	20	D27239	D27239 CELK016H5R
25		33.4	3.2	419	20	D27241	D27241 CELK004G8R
c 26		33.2	3.2	994	26	W42408	W42408 mc68a10.r1
27		33	3.2	430	37	AA716984	AA716984 vu61a02.r
28		32.8	3.2	378	44	AU042113	AU042113 AU042113
29		32.8	3.2	242	49	AV012937	AV012937 AV012937
30		32.6	3.2	622	37	AA667489	AA667489 vs27a09.r
c 31		32.4	3.1	292	37	AA685967	AA685967 EST108782
c 32		32.4	3.1	384	46	AI433947	AI433947 ti12a01.x
c 33		32.4	3.1	507	47	AI497448	AI497448 fb53f10.x
c 34		32.4	3.1	455	48	AI548191	AI548191 UI-R-C3-s
c 35		32.2	3.1	529	20	M62055	M62055 EST00111 Hi
c 36		32.2	3.1	499	23	H29966	H29966 yp44c05.r1
37		32.2	3.1	634	36	AA645428	AA645428 vq66c09.s
38		32.2	3.1	430	39	AA880213	AA880213 vx39b09.r
c 39		32	3.1	405	34	AA464948	AA464948 aa93h12.s
c 40		32	3.1	403	41	AA999281	AA999281 MCCE3087.
c 41		32	3.1	428	48	AI614040	AI614040 vg32h03.y
42		31.8	3.1	318	20	Z43599	Z43599 HSC1GA021 n
43		31.8	3.1	592	44	AU039090	AU039090 AU039090
44		31.8	3.1	823	49	AL048377	AL048377 DKFZp586E
c 45		31.8	3.1	747	51	AI723378	AI723378 hcgls21.T

ALIGNMENTS

RESULT 1
AA208468
LOCUS AA208468 323 bp mRNA EST 12-MAR-1997
DEFINITION mv85f02.r1 GuayWoodford Beier mouse kidney day 7 Mus musculus cDNA clone IMAGE:661851 5', mRNA sequence.
ACCESSION AA208468
NID g1804982
VERSION AA208468.1 GI:1804982

```
                /number=10
     exon       7658. .7784
                /number=10
     intron     7785. .7869
                /number=9
     exon       7870. .7954
                /number=9
     intron     7955. .8045
                /number=8
     exon       8046. .8109
                /number=8
     intron     8110. .8215
                /number=7
     exon       8216. .8277
                /number=7
     intron     8278. .8488
                /number=6
     exon       8489. .8559
                /number=6
     intron     8560. .8687
                /number=5
     exon       8688. .8783
                /number=5
     intron     8784. .8859
                /number=4
     exon       8860. .9008
                /number=4
     intron     9009. .9098
                /number=3
     exon       9099. .9168
                /number=3
     intron     9169. .9434
                /number=2
     exon       9435. .9475

  Query Match             15.4%;  Score 29.8;  DB 8;  Length 206606;
  Best Local Similarity   52.9%;  Pred. No. 29;
  Matches   64;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy      72 CACGAATCTTGCGGCGAATCTTTCCGCAGTCAGAGAAAGTGCCACAGCGACGCTGTCAGG 131
           || || | | |  | || || |||| || |    |||  |   | |     |  ||  |
Db   39129 CAAGAGTTTGGTAGAGATTCCTTCCTCATTTTCGGAACTTCGAAAACTAGAGACGTTGGT 39070

Qy     132 GGAAATTAAAGGCCCGCAACTGGAAGATTTTCCCGCGCTGATCAAACAGGCGAGTCTGGA 191
             |   ||  || |  | || |||| | |||| | || |||||   |||  ||| ||
Db   39069 AATACACAACTGCACAAAGCTAGAAGTTGTTCCAACTCTCATCAACTTGGCATCTCTTGA 39010

Qy     192 T 192
           |
Db   39009 T 39009


Search completed: October  5, 1999, 18:02:15
Job time: 11578 sec
```

```
  AUTHORS   EU Arabidopsis sequencing,project and ESSA.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JUN-1997) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, Project
            Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
            Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
            E-mail: michael.bevan@bbsrc.ac.uk
COMMENT     this fragment has,an overlap with ATFCA0 at the 5' end and an
            overlap
            with ATFCA2 at the 3' end.
FEATURES             Location/Qualifiers
     source          1. .206606
                     /organism="Arabidopsis thaliana"
                     /variety="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="4"
     CDS             complement(join(685. .744,829. .903,1053. .1232,1804. .1806))
                     /note="strong similarity to ribosomal protein L41 -
                     Candida maltosa"
                     /codon_start=1
                     /product="ribosomal protein"
                     /protein_id="CAB10211.1"
                     /db_xref="PID:e326883"
                     /db_xref="PID:g2244789"
                     /db_xref="GI:2244789"
                     /db_xref="SWISS-PROT:O23290"
                     /translation="MVNIPKTKNTYCKNKECKKHTLHKVTQYKKGKDSLAAQGKRRYD
                     RKQSGYGGQTKPVFHKKAKTTKKIVLRLQCQSCKHFSQRPIKRCKHFEIGGDKKGKGT
                     SLF"
     exon            688. .744
                     /number=4
     intron          745. .828
                     /number=3
     exon            829. .903
                     /number=3
     intron          904. .1052
                     /number=2
     exon            1053. .1232
                     /number=2
     intron          1233. .1803
                     /number=1
     exon            1804. .1806
                     /number=1
     CDS             join(2229. .2837,2914. .3069,3161. .3268,3358. .3468,
                     3569. .4957,5000. .5212,6340. .6428,6469. .6560,6706. .6749,
                     7316. .7384)
                     /note="similarity to kinesin osm-3 - Caenorhabditis
                     elegans"
                     /codon_start=1
                     /product="kinesin homolog"
                     /protein_id="CAB10212.1"
                     /db_xref="PID:e327448"
                     /db_xref="PID:g2244790"
                     /db_xref="GI:2244790"
                     /db_xref="SPTREMBL:O23291"
                     /translation="MAPTPSSSRSNQTQYTLIRTPQTKQRLNFHSKTPNPDGSKDPSP
                     PEHPVEVIGRIRDYPDRKEKSPSILQVNTDNQTVRVRADVGYRDFTLDGVSFSEQEGL
                     EEFYKKFIEERIKGVKVGNKCTIMMYGPTGAGKSHTMFGCGKEPGIVYRSLRDILGDS
                     DQDGVTFVQVTVLEVYNEEIYDLLSTNSSNNLGIGWPKGASTKVRLEVMGKKAKNASF
                     ISGTEAGKISKEIVKVEKRRIVKSTLCNERSSRSHCIIILDVPTVGGRLMLVDMAGSE
                     NIDQAGQTGFEAKMQTAKINQGNIALKRVVESIANGDSHVPFRDSKLTMLLQDSFEDD
                     KSKILMILCASPDPKEMHKTLCTLEYGAKAKCIVRGSHTPNKDKYGGDESASAVILGS
                     RIAAMDEFIIKLQSEKKQKEKERNEAQKQLKKKEEEVAALRSLLTQREACATNEEEIK
                     EKVNERTQLLKSELDKKLEECRRMAEEFVEMERRRMEERIVQQQEELEMMRRRLEEIE
                     VEFRRSNGGSVDETSGFAKRLRSLYSDDDPGMVKSMDLDMGDPEPVKQVWGAVSHQSS
                     NTISSNFTNLLQPKPSENMLTQMYPDRVCLSTVFEEEEVEEEEEKVIVEDKSICLITT
                     PMPSLNSEGLGKENCFNGADDKESASSRRLRIQNIFTLCGNQRELSQHSGQEEDQANI
                     ASPDKKDNQFFSITNKAEALAVEEAKENNISVDQRENGQLDIYVKWETAADNPRKLIT
                     TLRVTKDATLADLRKLIEIYLGSDNQAFTFLKLGVINLNQQAQKAFHFYLFEPCGAQV
                     AKEKESTVQATSLPLCNGHAYLATLRPGKSSQHKSLQPASPLPLNPIENMMEVTPISK
                     VTPNHQILLKTETERLGEADYINSLSLSHFLILFPRGESHQHRMLSFHSHQTSPSLSS
                     FPLLSRADADEPGLVLDITPLFEVVLEGPTLVLELAVVNDRHIAG"
     exon            2229. .2837
                     /number=1
     intron          2838. .2913
                     /number=1
     exon            2914. .3069
                     /number=2
     intron          3070. .3160
                     /number=2
     exon            3161. .3268
                     /number=3
     intron          3269. .3357
                     /number=3
     exon            3358. .3468
                     /number=4
     intron          3469. .3568
                     /number=4
     exon            3569. .4957
                     /number=5
     intron          4958. .4999
                     /number=5
     exon            5000. .5212
                     /number=6
     intron          5213. .6339
                     /number=6
     exon            6340. .6428
                     /number=7
     intron          6429. .6468
                     /number=7
     exon            6469. .6560
                     /number=8
     CDS             complement(join(6533. .6772,6888. .7024,7289. .7352,
                     7451. .7524,7658. .7784,7870. .7954,8046. .8109,8216. .8277,
                     8489. .8559,8688. .8783,8860. .9008,9099. .9168,9435. .9475,
                     9576. .9669))
                     /note="CKI1 mRNA for casein kinase I"
                     /codon_start=1
                     /product="casein kinase I"
                     /protein_id="CAB10213.1"
                     /db_xref="PID:e326884"
                     /db_xref="PID:g2244791"
                     /db_xref="GI:2244791"
                     /db_xref="SPTREMBL:Q39050"
                     /translation="MDRNQKMDHVIGGKFKLGRKLGSGSFGELYLGINIQTGEEVAVK
                     LEPVKTRHPQLQYESKIYMFLQGGTGVPHLKWFGVEGEYSCMVIDLLGPSLEDLFNYC
                     KRIFSLKSVLMLADQLICRVEYMHSRGFLHRDIKPDNFLMGLGRRANQVYIIDYGLAK
                     KYKDLQTQKHIPYRENKNLTGTARYASVNTHLGIEQSRRDDLESLGYVLMYFLRGSLP
                     WQGLKAGTKKQKYDKISEKKMLTSVETLCKSYPSEFTSYFHYCRSLRFEDKPDYSYLR
                     RLFRDLFIREGYQLDYVFDWTISKYPQIGSSSRPRPTPRPALDPPGPPAERAEKPTVG
                     QDLRGRFTGAIEAFTRRNVSSQGALGDRSRHRSSDDIPSSAKEVHESRNGSTSKRGVI
                     SSTRPGSSAEPSENHSSRLFSSGSRHATTQRVPQSYESAAAARPGHEDAIRNFELLTI
                     GSGKKRK"
     exon            6536. .6772
                     /number=14
     intron          6561. .6705
                     /number=8
     exon            6706. .6749
                     /number=9
     intron          6750. .7315
                     /number=9
     intron          6773. .6887
                     /number=13
     exon            6888. .7024
                     /number=13
     intron          7025. .7288
                     /number=12
     exon            7289. .7352
                     /number=12
     exon            7316. .7384
                     /number=10
     intron          7353. .7450
                     /number=11
     exon            7451. .7524
                     /number=11
     intron          7525. .7657
```

```
                TGTDTSLESFRDLVSNHNWDDFGNNVPLDDDEEGSEDPLRQLALENAPETSNYDGAED
                LLFDNEQQYPENGFDVPDPNYLPLADEEIFPHSPALRTPSPSDPNLV"
     gene       13950. .15270
                /gene="Y53C10A.1"
     CDS        join(13950. .13973,14303. .14448,14498. .14870,14922. .15155,
                15205. .15270)
                /gene="Y53C10A.1"
                /note="predicted using Genefinder"
                /codon_start=1
                /protein_id="CAA22135.1"
                /db_xref="PID:e1356170"
                /db_xref="PID:g3947648"
                /db_xref="GI:3947648"
                /translation="MSPELEKLNKKTGEVRLDDKQRIIEYASKNGKIVLKGDHSASAQ
                RRRGLKRRRDGEGTSAPHAEDSSDDDVEFVGLVPCVPKQEPAARIDFDEYFNRDVSFN
                PEMQQVPPAIDRNCENPTFDETDTKNNVQLVGATRNGSNSDGISLHMLVKRLKEMLIY
                LGSDYTHLMGKIYYLLAIRDEKLNAEDVIAFLKTFLIINTARKFRLSTESIGYTKSSK
                KRATEFLSILKFVLLNLESPQMQEFQNEIENKIENIGAVEKVILVVDAQEAIQQILST
                FSSS"
     gene       complement(15701. .16222)
                /gene="Y53C10A.2"
     CDS        complement(join(15701. .15727,15787. .15882,15926. .16222))
                /gene="Y53C10A.2"
                /note="predicted using Genefinder"
                /codon_start=1
                /protein_id="CAA22136.1"
                /db_xref="PID:e1356171"
                /db_xref="PID:g3947649"
                /db_xref="GI:3947649"
                /translation="MIQDIKRSSAPAAPAPVTVLKMWPNEMRSQSLEENLDAKLEQFK
                KEHSPVKERGNCSKTRKISENQPPPSFEVVMAKIRGIQESLKNMEDQMNEEERGKEMP
                DSTENSFDFSKIWQYFDQTRIEEIRALLKKRQFPDSP"
     gene       complement(16631. .19476)
                /gene="Y53C10A.3"
     CDS        complement(join(16631. .16713,16865. .17008,17196. .17409,
                17458. .17542,18991. .19076,19129. .19476))
                /gene="Y53C10A.3"
                /note="predicted using Genefinder"
                /codon_start=1
                /protein_id="CAA22137.1"
                /db_xref="PID:e1356172"
                /db_xref="PID:g3947650"
                /db_xref="GI:3947650"
                /translation="MVKLYIIIAFIIVSISTSSPKPTKCGTIDDDKSKIPNLLMEPSK
                YFVFNGPFDKLSMDYFILRHPEFQRIAYNVTLKQSKDSKIFLRDPNGIIEENATEKLV
                HIGFEPFNYTENNKEFSKDKIRIQWMNVPKDASPDSIESCLDTFSVRKMCAYQGNPSI
                SRLLQMQQDQQKMIEQTKNMSIDEQNARTVLINLWHVLESRMFQSIAYWDNESFHISD
                PYLFERHVLKLFKTNDINDIFKVLFMCGFGKITHVSHGCHVHLEFSHPFFKKGRFDLL
                SQIKIQVGDSDEYEKKYPHGLHSGLLYSASSASSDKSYHTNSY"
     tRNA       complement(24870. .24941)
                /gene="Y53C10A.t3"
                /note="CTC Glu E-tRNA; predicted using tRNAscan-SE-1.11;
                preliminary prediction; similar to tRNA-Glu"
     gene       complement(24870. .24941)
                /gene="Y53C10A.t3"
     tRNA       25671. .25742
                /gene="Y53C10A.t1"
                /note="CTC Glu E-tRNA; predicted using tRNAscan-SE-1.11;
                preliminary prediction; similar to tRNA-Glu"
     gene       25671. .25742
                /gene="Y53C10A.t1"
     tRNA       26443. .26514
                /gene="Y53C10A.t2"
                /note="CTC Glu E-tRNA; predicted using tRNAscan-SE-1.11;
                preliminary prediction; similar to tRNA-Glu"
     gene       26443. .26514
                /gene="Y53C10A.t2"
     gene       27186. .33117
                /gene="Y53C10A.4"
     CDS        join(27186. .27230,28228. .28355,28423. .28545,28599. .28713,
                28755. .28863,29522. .29723,29984. .30708,30965. .31316,
                31576. .31879,32076. .32424,32646. .32751,32949. .33117)
                /gene="Y53C10A.4"
                /note="predicted using Genefinder; similar to RhoGAP
                domain; cDNA EST yk472h10.5 comes from this gene; cDNA EST
                yk477f5.5 comes from this gene; cDNA EST yk474f7.5 comes
                from this gene; cDNA EST yk474f7.3 comes from this gene"
                /codon_start=1
                /protein_id="CAA22138.1"
                /db_xref="PID:e1356173"
                /db_xref="PID:g3947651"
                /db_xref="GI:3947651"
                /translation="MISSNSSSTSLNVETDYPTNGKKPLKEGSVQLTSLSTLLSHHRY
                FFLYHETLIISKQKGACSYKLKEKLRLEKVWIASSNTADSFLIGWPFTNYLVHFRNTE
                EKDEWFELLSCCVQQCLRPLFTTISMEINVRGRKQTIRRRVDNGKKSGEIVVDTAGDL
                GLPHTSYQLTLSFGDNSHRQLQGPENVYVAIMSEIERQGTRLSESQKQALDTCPIANC
                RLILSTIKSTSGPSAKQLVNSIKKKVLTRTDSRSIFGKGLNGPTPPQPIMTIVDHLRM
                DGFDAEGIFRKSPKQSTFKELKCELDKGVVPDFHKYNTHVLASILKEYLRSIPGKILL
                SGNYELWMREIADEPNTEKKVSCCRALLSHLPTSHSILLANVLKLLNKISNSPSSKMN
                ASSLSVCLAPSFLESPDPMEGGKKIPPLIEFLISHAAQVMPGFSTDNVFAIMTSPMVD
                HNANMVCSPAPSSSHSDDGGSSQKTPIIEELNDADDATSQSSLAESDLPDEPRGTSAH
                LETISEISEPPTNDERLVFSSSEDEDDDDEDELRVYRPKSSTSTSSCFARTSFVSETL
                KTSADDGDKGLNKETPRSPCLKRIHFQRTEAQHRNKEGNIKLDFEALKFHRPSVRKSS
                LEEVARKNSAEFFHHSREPSDAAVTPTPPKNGSTMRKESREEATQTADCGLPHISSND
                DSDSRFPAFKSSSVAMLTEPYKSSSSSSSSTSSIPSDPAPRPYFLKEEYTSLANKPTV
                QVAPFSQEPSLKTINDRLNALRQNSILAQEVPQEPVMTSSRLRSEATKSLTNSPIMPR
                SYGGSLMQKDREEHSFVADDFYKASSRKPELQRSSSQRGLTSMNNNNNSTKQIYGARP
                SSLKTPRRFAGAGGVEACVGMDPLEINWSVSQLKTLFQDTKAPSIDTTYNMN"
     gene       46461. .50043
                /gene="Y53C10A.5"
     CDS        join(46461. .46552,46704. .46882,47412. .47510,47583. .47911,
                48897. .48989,49043. .49118,49952. .50043)
                /gene="Y53C10A.5"
                /note="predicted using Genefinder"
                /codon_start=1
                /protein_id="CAA22139.1"
                /db_xref="PID:e1356174"
                /db_xref="PID:g3947652"
                /db_xref="GI:3947652"
                /translation="MTSSSAASSSTTTSSTMMPDENECLKKEEERFKSPDPAPTLDEE
```

```
  Query Match              15.6%;  Score 30.2;  DB 36;  Length 78153;
  Best Local Similarity    55.1%;  Pred. No. 20;
  Matches   59;  Conservative    0;  Mismatches   48;  Indels    0;  Gaps    0;

Qy      80 TTGCGGCGAATCTTTCCGCAGTCAGAGAAAGTGCCACAGCGACGCTGTCAGGGGAAATTA 139
           || | |  ||  ||||  ||    |  ||  |||||   || || | |    | || ||
Db   29089 TTTCCGAAAAATTTTCGACAAATTGGCAATTTGCCAGTTCGCCGATTTGCCTGAAATTTC 29148

Qy     140 AAGGCCCGCAACTGGAAGATTTTCCCGCGCTGATCAAACAGGCGAGT 186
           ||  || |||| | |  ||||| || |   |  |||||   || | |
Db   29149 AATTCCGGCAATTTGTCGATTTGCCAGAAATTTTCAAATTTGCAATT 29195


RESULT  15
ATFCA1/c
LOCUS       ATFCA1     206606 bp    DNA             PLN       10-FEB-1999
DEFINITION  Arabidopsis thaliana DNA chromosome 4, ESSA I contig fragment No.
            1.
ACCESSION   Z97336
NID         g2244788
VERSION     Z97336.1  GI:2244788
KEYWORDS    .
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
            Rosidae; Capparales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 206606)
  AUTHORS   Bevan,M., Stiekema,W., Murphy,G., Wambutt,R., Pohl,T., Terryn,N.,
            Kreis,M., Kavanagh,T., Entian,K.D., Rieger,M., James,R.,
            Puigdomenech,P., Hatzopoulos,P., Obermaier,B., Duesterhoft,A.,
            Jones,J., Palme,K., Ansorge,W., Delseny,M., Bancroft,I.,
            Mewes,H.W., Schueller,C. and Chalwatzis,N.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 206606)
```

```
            *     3761     3890: contig of 130 bp in length
            *                    gap of unknown length
            *     3891     4876: contig of 986 bp in length
            *                    gap of unknown length
            *     4877     5825: contig of 949 bp in length
            *                    gap of unknown length
            *     5826     6423: contig of 598 bp in length
            *                    gap of unknown length
            *     6424     7586: contig of 1163 bp in length
            *                    gap of unknown length
            *     7587     8570: contig of 984 bp in length
            *                    gap of unknown length
            *     8571     9584: contig of 1014 bp in length
            *                    gap of unknown length
            *     9585    10993: contig of 1409 bp in length
            *                    gap of unknown length
            *    10994    12180: contig of 1187 bp in length
            *                    gap of unknown length
            *    12181    14564: contig of 2384 bp in length
            *                    gap of unknown length
            *    14565    19619: contig of 5055 bp in length
            *                    gap of unknown length
            *    19620    29845: contig of 10226 bp in length
            *                    gap of unknown length
            *    29846    41025: contig of 11180 bp in length
            *                    gap of unknown length
            *    41026    75786: contig of 34761 bp in length
            *                    gap of unknown length
            *    75787   134036: contig of 58250 bp in length
            *                    gap of unknown length
            *   134037   175591: contig of 41555 bp in length.
FEATURES             Location/Qualifiers
     source          1. .175591
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="35p16"
                     /chromosome="16"
BASE COUNT    44292 a  43232 c  44714 g  43331 t     22 others
ORIGIN


  Query Match             15.6%;  Score 30.2;  DB 35;  Length 175591;
  Best Local Similarity   53.9%;  Pred. No. 22;
  Matches   62;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy      10 AGTGTAAGGACTCAGCCCCCCGTCATAATGCCAGGTATGCAGACCGAGATCAAAACGCAG 69
           || | | ||| | |||  |  |  | |||         || |   | ||| || || |
Db    4267 AGCGAACGGATTGAGCAACAGGAGACAATATTGCAGCAGCTGGTGGCGATGAACACCCGA 4208

Qy      70 GCCACGAATCTTGCGGCGAATCTTTCCGCAGTCAGAGAAAGTGCCACAGCGACGC 124
             | |||    ||| || || ||   ||||   || ||  ||||||| | ||||
Db    4207 CTGAAGAACGCCGCGCCGGATATTATTGCAGCGCGAAAATCTGCCACAACAACGC 4153


RESULT  14
CEY53C10A
LOCUS       CEY53C10A   78153 bp    DNA             INV       18-DEC-1998
DEFINITION  Caenorhabditis elegans cosmid Y53C10A, complete sequence.
ACCESSION   AL033536
NID         g3881162
VERSION     AL033536.1  GI:3881162
KEYWORDS    HTG.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 78153)
  AUTHORS   White,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-NOV-1998) Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
REFERENCE   2  (bases 1 to 78153)
  AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
  TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
  JOURNAL   Nature 368 (6466), 32-38 (1994)
  MEDLINE   94150718
COMMENT     Coding sequences below are predicted from computer analysis, using
            predictions from Genefinder (P. Green, U. Washington), and other
            available information.
            For a graphical representation of this sequence and its analysis
            see:-
            http://webace.sanger.ac.uk/cgi-
            bin/display?db=wormace&class=Sequence &object=Y53C10A
            Current sequence finishing criteria for the C. elegans genome
            sequencing consortium are that all bases are either sequenced
            unambiguously on both strands, or on a single strand with both a
            dye primer and dye terminator reaction, from distinct subclones.
            Exceptions are indicated by an explicit note.
            IMPORTANT:  This sequence is NOT necessarily the entire insert of
            the specified clone.  It may be shorter because we only sequence
            overlapping sections once, or longer because we arrange for a small
            overlap between neighbouring submissions.
            IMPORTANT: This sequence is not the entire insert of clone Y53C10A.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we arrange for a small overlap between
            neighbouring submissions.
            The true left end of clone Y47H10 is at 78045 in this sequence. The
            true right end of clone R06C1 is at 107 in this sequence. The start
            of this sequence (1. .107) overlaps with the end of sequence Z81106.
            The end of this sequence overlaps with the start of sequence
            Z95311.
FEATURES             Location/Qualifiers
     source          1. .78153
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /clone="Y53C10A"
     gene            complement(1476. .9641)
                     /gene="Y53C10A.12"
     CDS             complement(join(1476. .1568,3002. .3484,4354. .4942,
                     5915. .6049,6446. .6579,6790. .6923,7596. .7815,9414. .9641))
                     /gene="Y53C10A.12"
                     /note="similar to HSF-type DNA-binding domain; cDNA EST
                     yk485h11.5 comes from this gene; cDNA EST yk208g2.5 comes
                     from this gene; cDNA EST yk347g12.5 comes from this gene;
                     cDNA EST EMBL:C09726 comes from this gene; cDNA EST
                     yk446f8.5 comes from this gene; cDNA EST EMBL:D27715 comes
                     from this gene; cDNA EST EMBL:C08108 comes from this gene;
                     cDNA EST yk295c5.5 comes from this gene; cDNA EST
                     EMBL:D27716 comes from this gene; cDNA EST yk446f8.3 comes
                     from this gene; cDNA EST yk295c5.3 comes from this gene"
                     /codon_start=1
                     /protein_id="CAA22146.1"
                     /db_xref="PID:e1356181"
                     /db_xref="PID:g3947659"
                     /db_xref="GI:3947659"
                     /translation="MQPTGNQIQQNQQQQQQLIMRVPKQEVSVSGAARRYVQQAPPNR
                     PPRQNHQNGAIGGKKSSVTIQEVPNNAYLETLNKSGNNKVDDDKLPVFLIKLWNIVED
                     PNLQSIVHWDDSGASFHISDPYLFGRNVLPHFFKHNNMNSMVRQLNMYGFRKMTPLSQ
                     GGLTRTESDQDHLEFSHPCFVQGRPELLSQIKRKQSARTVEDKQVNEQTQQNLEVVMA
                     EMRAMREKAKNMEDKMNKLTKENRDMWTQMGSMRQQHARQQQYFKKLLHFLVSVMQPG
                     LSKRVAKRGVLEIDFCAANGTAGPNSKRARMNSEEGPYKDVCDLLESLQRETQEPFSR
                     RFTNNEGPLISEVTDEFGNSPVGRGSAQDLFGDTFGAQSSRYSDGGATSSREQSPHPI
                     ISQPQSNSAGAHGANEQKPDDMYMGSGPLTHENIHRGISALKRDYQGASPASGGPSTS
                     SSAPSGAGAGARMAQKRAAPYKNATRQMAQPQQDYSGGFVNNYSGFMPSDPSMIPYQP
                     SHQYLQPHQKLMAIEDQHHPTTSTSSTNADPHQNLYSPTLGLSPSFDRQLSQELQEYF
```

```
     promoter        complement(4895. .4927)
                     /note="factor Sigma70; promoter speCp3; documented +1 at
                     3107277"
     promoter        complement(4912. .4942)
                     /note="factor Sigma70; promoter speCp2; documented +1 at
                     3107292"
     promoter        complement(4930. .4962)
                     /note="factor Sigma70; promoter speCp1; documented +1 at
                     3107304"
     promoter        5010. .5040
                     /note="factor Sigma70; predicted +1 start at 3107434"
     gene            5183. .5890
                     /gene="yqgA"
                     /note="b2966"
     CDS             5183. .5890
                     /gene="yqgA"
                     /function="putative transport; Not classified"
                     /note="o235; This 235 aa ORF is 23 pct identical (10 gaps)
                     to 156 residues of an approx. 824 aa protein SNF3_YEAST
                     SW: P10870"
                     /codon_start=1
                     /transl_table=11
                     /product="putative transport protein"
                     /protein_id="AAC76003.1"
                     /db_xref="PID:g1789338"
                     /db_xref="GI:1789338"
                     /translation="MVIGPFINASAVLLGGVLGALLSQRLPERIRVSMTSIFGLASLG
                     IGILLVVKCANLPAMVLATLLGALIGEICLLEKGVNTAVAKAQNLFRHSRKKPAHESF
                     IQNYVAIIVLFCASGTGIFGAMNEGMTGDPSILIAKSFLDFFTAMIFACSLGIAVSVI
                     SIPLLIIQLTLAWAAALILPLTTPSMMADFSAVGGLLLLATGLRICGIKMFPVVNMLP
                     ALLLAMPLSAAWTAWFA"
     tRNA            5996. .6071
                     /gene="pheV"
                     /note="anticodon: GAA; CG Site No. 18058"
                     /product="tRNA-Phe"
                     /function="RNA; tRNA"
     gene            5996. .6071
                     /gene="pheV"
                     /note="b2967"
     gene            complement(6220. .6756)
                     /gene="yghD"
                     /note="b2968"
     CDS             complement(6220. .6756)
                     /gene="yghD"
                     /function="putative transport; Not classified"
                     /note="f178; This 178 aa ORF is 32 pct identical (5 gaps)
                     to 146 residues of an approx. 168 aa protein GSPM_VIBCH
                     SW: P41851"
                     /codon_start=1
                     /transl_table=11
                     /product="putative secretion pathway protein"
                     /protein_id="AAC76004.1"

  Query Match             15.6%;  Score 30.2;  DB 2;  Length 10170;
  Best Local Similarity   53.9%;  Pred. No. 17;
  Matches   62;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy      10 AGTGTAAGGACTCAGCCCCCCGTCATAATGCCAGGTATGCAGACCGAGATCAAAACGCAG 69
           || | | ||| | |||  |  |  | |||         || |   | ||| || || |
Db    6576 AGCGAACGGATTGAGCAACAGGAGACAATATTGCAGCAGCTGGTGGCGATGAACACCCGA 6517

Qy      70 GCCACGAATCTTGCGGCGAATCTTTCCGCAGTCAGAGAAAGTGCCACAGCGACGC 124
             | |||    ||| || || ||   ||||   || ||  ||||||| | ||||
Db    6516 CTGAAGAACGCCGCGCCGGATATTATTGCAGCGCGAAAATCTGCCACAACAACGC 6462


RESULT  12
CEY53C10_2
WPCOMMENT
  Sequence split into 5 fragments  LOCUS CEY53C10 Accession Z93340
     Fragment Name      Begin     End
•    CEY53C10_0             1    110000
     CEY53C10_1        100001    210000
     CEY53C10_2        200001    310000
     CEY53C10_3        300001    410000
     CEY53C10_4        400001    413826
Continuation (3 of 5) of CEY53C10 from base 200001 (Z93340 Caenorhabditis elegans chr


  Query Match             15.6%;  Score 30.2;  DB 34;  Length 110000;
  Best Local Similarity   55.1%;  Pred. No. 21;
  Matches   59;  Conservative    0;  Mismatches   48;  Indels    0;  Gaps    0;

Qy      80 TTGCGGCGAATCTTTCCGCAGTCAGAGAAAGTGCCACAGCGACGCTGTCAGGGGAAATTA 139
           || | |  ||  ||||  ||     |  ||  |||||   || || | |    | || ||
Db   72557 TTTCCGAAAAATTTTCGACAAATTGGCAATTTGCCAGTTCGCCGATTTGCCTGAAATTTC 72616

Qy     140 AAGGCCCGCAACTGGAAGATTTTCCCGCGCTGATCAAACAGGCGAGT 186
           ||  || |||| | |  ||||| || |   |  |||||   || | |
Db   72617 AATTCCGGCAATTTGTCGATTTGCCAGAAATTTTCAAATTTGCAATT 72663


RESULT  13
AC007606/c
LOCUS       AC007606   175591 bp    DNA             HTG       20-MAY-1999
DEFINITION  Homo sapiens chromosome 16 clone 35p16, WORKING DRAFT SEQUENCE, 21
            unordered pieces.
ACCESSION   AC007606
NID         g4878060
VERSION     AC007606.1  GI:4878060
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 175591)
  AUTHORS   Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E., Robinson,D.,
            Jones,M., Buckingham,J., Chasteen,L., Thompson,S., Goodwin,L.,
            Bryant,J., Tesmer,J., Meincke,L., Longmire,J., White,S., Tatum,O.,
            Campbell,C., Fawcett,J., Maltbie,M., Bussod,M., Sutherland,R.,
            Han,C. and Deaven,L.
  TITLE     Sequencing of Human Chromosome 16q12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 175591)
  AUTHORS   Ricke,D.O.
  TITLE     Large Scale Sequence Analysis and Annotation with the Sequence
            Comparison Analysis (SCAN) System
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 175591)
  AUTHORS   Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E., Robinson,D.,
            Jones,M., Buckingham,J., Chasteen,L., Thompson,S., Goodwin,L.,
            Bryant,J., Tesmer,J., Meincke,L., Longmire,J., White,S., Tatum,O.,
            Campbell,C., Fawcett,J., Maltbie,M., Bussod,M., Sutherland,R.,
            McMurry,K., Han,C. and Deaven,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-MAY-1999) Center for Human Genome Studies, DOE Joint
            Genome Institute, Los Alamos National Laboratory, MS M888, Los
            Alamos, NM 87545, USA
COMMENT     * NOTE: This is a 'working draft' sequence. It currently
            * consists of 21 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      532: contig of 532 bp in length
            *                    gap of unknown length
            *      533     1114: contig of 582 bp in length
            *                    gap of unknown length
            *     1115     1773: contig of 659 bp in length
            *                    gap of unknown length
            *     1774     2783: contig of 1010 bp in length
            *                    gap of unknown length
            *     2784     3760: contig of 977 bp in length
            *                    gap of unknown length
```

genome.
ACCESSION AE000379 U00096
NID g2367179
VERSION AE000379.1 GI:2367179
KEYWORDS .
SOURCE Escherichia coli.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 10170)
AUTHORS Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V., Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F., Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J., Mau,B. and Shao,Y.
TITLE The complete genome sequence of Escherichia coli K-12
JOURNAL Science 277 (5331), 1453-1474 (1997)
MEDLINE 97426617
REFERENCE 2 (bases 1 to 10170)
AUTHORS Blattner,F.R.
TITLE Direct Submission
JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA. Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
REFERENCE 3 (bases 1 to 10170)
AUTHORS Blattner,F.R.
TITLE Direct Submission
JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA. Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
REFERENCE 4 (bases 1 to 10170)
AUTHORS Plunkett,G. III.
TITLE Direct Submission
JOURNAL Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT On Sep 9, 1997 this sequence version replaced gi:1789333. This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants HG00301 and HG01428 (from the Human Genome Project and NCHGR). The entire sequence was independently determined from E. coli K-12 strain MG1655. Predicted open reading frames were determined using GeneMark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (http://cgsc.biology.yale.edu). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (http://www.genetics.wisc.edu). *** The E. coli K-12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.
FEATURES Location/Qualifiers
source 1. .10170
/organism="Escherichia coli"
/strain="K-12"
/sub_strain="MG1655"
/db_xref="taxon:562"
gene 61. .1143
/gene="mltC"
/note="b2963"
CDS 61. .1143
/gene="mltC"
/function="putative enzyme; Murein sacculus, peptidoglycan"
/note="o360; formerly designated yggZ"
/codon_start=1
/transl_table=11
/product="membrane-bound lytic murein transglycosylase C"
/protein_id="AAC76000.1"
/db_xref="PID:g2367180"
/db_xref="GI:2367180"
/translation="MMKKYLALALIAPLLISCSTTKKGDTYNEAWVKDINGFDILMGQ
FAHNIENIWGFKEVVIAGPKDYVKYTDQYQTRSHINFDDGTITIETIAGTEPAAHLRR
AIIKTLLMGDDPSSVDLYSDVDDITISKEPFLYGQVVDNTGQPIRWEGRASNFADYLL
KNRLKSRSNGLRIIYSVTINMVPNHLDKRAHKYLGMVRQASRKYGVDESLILAIMQTE
SSFNPYAVSRSDALGLMQVVQHTAGKDVFRSQGKSGTPSRSFLFDPASNIDTGTAYLA
MLNNVYLGGIDNPTSRRYAVITAYNGGAGSVLRVFSNDKIQAANIINTMTPGDVYQTL
TTRHPSAESRRYLYKVNTAQKSYRRR"
promoter 1205. .1231
/note="factor Sigma70; predicted +1 start at 3103625"
gene 1297. .2601
/gene="nupG"
/note="b2964"
CDS 1297. .2601
/gene="nupG"
/function="transport; Transport of small molecules: Nucleosides, purines, pyrimidines"
/note="o434; 100 pct identical to NUPG_ECOLI SW: P09452 but has 16 additional N-terminal residues; CG Site No. 442"
/codon_start=1
/transl_table=11
/product="transport of nucleosides, permease protein"
/protein_id="AAC76001.1"
/db_xref="PID:g1789336"
/db_xref="GI:1789336"
/translation="MCFFKHSSASRCEEINMNLKLQLKILSFLQFCLWGSWLTTLGSY
MFVTLKFDGASIGAVYSSLGIAAVFMPALLGIVADKWLSAKWVYAICHTIGAITLFMA
AQVTTPEAMFLVILINSFAYMPTLGLINTISYYRLQNAGMDIVTDFPPIRIWGTIGFI
MAMWVVSLSGFELSHMQLYIGAALSAILVLFTLTLPHIPVAKQQANQSWTTLLGLDAF
ALFKNKRMAIFFIFSMLLGAELQITNMFGNTFLHSFDKDPMFASSFIVQHASIIMSIS
QISETLFILTIPFFLSRYGIKNVMMISIVAWILRFALFAYGDPTPFGTVLLVLSMIVY
GCAFDFFNISGSVFVEKEVSPAIRASAQGMFLMMTNGFGCILGGIVSGKVVEMYTQNG
ITDWQTVWLIFAGYSVVLAFAFMAMFKYKHVRVPTGTQTVSH"
gene complement(2651. .4846)
/gene="speC"
/note="b2965"
CDS complement(2651. .4846)
/gene="speC"
/EC_number="4.1.1.17"
/function="enzyme; Central intermediary metabolism: Polyamine biosynthesis"
/note="f731; 100 pct identical to DCOR_ECOLI SW: P21169; CG Site No. 159"
/codon_start=1
/transl_table=11
/product="ornithine decarboxylase isozyme"
/protein_id="AAC76002.1"
/db_xref="PID:g1789337"
/db_xref="GI:1789337"
/translation="MGQGFPPCPVFLLPRNGFALMKSMNIAASSELVSRLSSHRRVVA
LGDTDFTDVAAVVITAADSRSGILALLKRTGFHLPVFLYSEHAVELPAGVTAVINGNE
QQWLELESAACQYEENLLPPFYDTLTQYVEMGNSTFACPGHQHGAFFKKHPAGRHFYD
FFGENVFRADMCNADVKLGDLLIHEGSAKDAQKFAAKVFHADKTYFVLNGTSAANKVV
TNALLTRGDLVLFDRNNHKSNHHGALIQAGATPVYLEASRNPFGFIGGIDAHCFNEEY
LRQQIRDVAPEKADLPRPYRLAIIQLGTYDGTVYNARQVIDTVGHLCDYILFDSAWVG
YEQFIPMMADSSPLLLELNENDPGIFVTQSVHKQQAGFSQTSQIHKKDNHIRGQARFC
PHKRLNNAFMLHASTSPFYPLFAALDVNAKIHEGESGRRLWAECVEIGIEARKAILAR
CKLFRPFIPPVVDGKLWQDYPTSVLASDRRFFSFEPGAKWHGFEGYAADQYFVDPCKL
LLTTPGIDAETGEYSDFGVPATILAHYLRENGIVPEKCDLNSILFLLTPAESHEKLAQ
LVAMLAQFEQHIEDDSPLVEVLPSVYNKYPVRYRDYTLRQLCQEMHDLYVSFDVKDLQ
KAMFRQQSFPSVVMNPQDAHSAYIRGDVELVRIRDAEGRIAAEGALPYPPGVLCVVPG
EVWGGAVQRYFLALEEGVNLLPGFSPELQGVYSETDADGVKRLYGYVLK"

```
                     VEVADKAGVLTRAEAAEINLRSDILNAVGITLDETTTRENVMQLFNVLLGDNHGLDID
                     TLDKDVAHDSRSIQPAMLRDDEILTHPVFNRYHSETEMMRYMHSLERKDLALNQAMIP
                     LGSCTMKLNAAAEMIPITWPEFAELHPFCPPEQAEGYQQMIAQLADWLVKLTGYDAVC
                     MQPNSGAQGEYAGLLAIRHYHESRNEGHRDICLIPASAHGTNPASAHMAGMQVVVVAC
                     DKNGNIDLTDLRAKAEQAGDNLSCIMVTYPSTHGVYEETIREVCEVVHQFGGQVYLDG
                     ANMNAQVGITSPGFIGADVSHLNLHKTFCIPHGGGGPGMGPIGVKAHLAPFVPGHSVV
                     QIEGMLTRQGAVSAAPFGSASILPISWMYIRMMGAEGLKKASQVAILNANYIASRLQD
                     AFPVLYTGRDGRVAHECILDIRPLKEETGISELDIAKRLIDYGFHAPTMSFPVAGTLM
                     VEPTESESKVELDRFIDAMLAIRAEIDQVKAGVWPLEDNPLVNAPHIQSELVAEWAHP
                     YSREVAVFPAGVADKYWPTVKRLDDVYGDRNLFCSCVPISEYQ"
gene                 complement(3551. .3940)
                     /gene="gcvH"
CDS                  complement(3551. .3940)
                     /gene="gcvH"
                     /note="ORF_f129"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA69072.1"
                     /db_xref="PID:g882434"
                     /db_xref="GI:882434"
                     /translation="MSNVPAELKYSKEHEWLRKEADGTYTVGITEHAQELLGDMVFVD
                     LPEVGATVSAGDDCAVAESVKAASDIYAPVSGEIVAVNDALSDSPELVNSEPYAGGWI
                     FKIKASDESELESLLDATAYEALLEDE"
misc_feature         complement(3809. .5631)
                     /note="similar to GenBank Accession Number M97263
                     (ECOTPRO)"
gene                 complement(3964. .5058)
                     /gene="gcvT"
CDS                  complement(3964. .5058)
                     /gene="gcvT"
                     /note="ORF_f364"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA69073.1"
                     /db_xref="PID:g882435"
                     /db_xref="GI:882435"
                     /translation="MAQQTPLYEQHTLCGARMVDFHGWMMPLHYGSQIDEHHAVRTDA
                     GMFDVSHMTIVDLRGSRTREFLRYLLANDVAKLTKSGKALYSGMLNASGGVIDDLIVY
                     YFTEDFFRLVVNSATREKDLSWITQHAEPFGIEITVRDDLSMIAVQGPNAQAKAATLF
                     NDAQRQAVEGMKPFFGVQAGDLFIATTGYTGEAGYEIALPNEKAADFWRALVEAGVKP
                     CGLGARDTLRLEAGMNLYGQEMDETISPLAANMGWTIAWEPADRDFIGREALEVQREH
                     GTEKLVGLVMTEKGVLRNELPVRFTDAQGNQHEGIITSGTFSPTLGYSIALARVPEGI
                     GETAIVQIRNREMPVKVTKPVFVRNGKAVA"
misc_feature         complement(5370. .10317)
                     /note="similar to GenBank Accession Number D90281
                     (ECOPUV)"
gene                 complement(5506. .6708)
                     /gene="visC"
CDS                  complement(5506. .6708)
                     /gene="visC"
                     /note="ORF_f400"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA69074.1"
                     /db_xref="PID:g882436"
                     /db_xref="GI:882436"
                     /translation="MQSVDVAIVGGGMVGLAVACGLQGSGLRVAVLEQRVQEPLAANA
                     PPQLRVSAINAASEKLLTRLGVWQDILSRRASCYHGMEVWDKDSFGHISFDDQSMGYS
                     HLGHIVENSVIHYALWNKAHQSSDITLLAPAELQQVAWGENETFLTLKDGSMLTARLV
                     IGADGANSWLRNKADIPLTFWDYQHHALVATIRTEEPHDAVARQVFHGEGILAFLPLS
                     DPHLCSIVWSLSPEEAQRMQQASEDEFNRALNIAFDNRLGLCKVESARQVFPLTGRYA
                     RQFASHRLALVGDAAHTIHPLAGQGVNLGFMDAAELIAELKRLHRQGKDIGQYIYLRR
                     YERSRKHSAALMLAGMQGFRDLFSGTNPAKKLLRDIGLKLADTLPGVKPQLIRQAMGL
                     NDLPEWLR"
gene                 complement(6731. .7909)
                     /gene="ubiH"
CDS                  complement(6731. .7909)
                     /gene="ubiH"
                     /note="ORF_f392"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA69075.1"
                     /db_xref="PID:g882437"
                     /db_xref="GI:882437"
                     /translation="MSVIIVGGGMAGATLALAISRLSHGALPVHLIEATAPESHAHPG
                     FDGRAIALAAGTCQQLARIGVWQSLADCATAITTVHVSDRGHAGFVTLAAEDYQLAAL
                     GQVVELHNVGQRLFALLRKAPGVTLHCPDRVANVARTQSHVEVTLESGETLTGRVLVA
                     ADGTHSALATACGVDWQQEPYEQLAVIANVATSVAHEGRAFERFTQHGPLAMLPMSDG
                     RCSLVWCHPLERREEVLSWSDEKFCRELQSAFGWRLGKITHAGKRSAYPLALTHAARS
                     ITHRTVLVGNAAQTLHPIAGQGFNLGMRDVMSLAETLTQAQERGEDMGDYGVLCRYQQ
                     RRQSDREATIGVTDSLVHLFANRWAPLVVGRNIGLMTMELFTPARDVLAQRTLGWVAR
                     "
misc_feature         complement(7749. .9599)
                     /note="similar to GenBank Accession Number D00398
                     (ECOAPP2)"
gene                 complement(7906. .9231)
                     /gene="pepP"
CDS                  complement(7906. .9231)
                     /gene="pepP"
                     /note="ORF_f441; third start codon"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA69076.1"
                     /db_xref="PID:g882438"
                     /db_xref="GI:882438"
                     /translation="MSEISRQEFQRRRQALVEQMQPGSAALIFAAPEVTRSADSEYPY
                     RQNSDFWYFTGFNEPEAVLVLIKSDDTHNHSVLFNRVRDLTAEIWFGRRLGQDAAPEK
                     LGVDRALAFSEINQQLYQLLNGLDVVYHAQGEYAYADVIVNSALEKLRKGSRQNLTAP
                     ATMIDWRPVVHEMRLFKSPEEIAVLRRAGEITAMAHTRAMEKCRPGMFEYHLEGEIHH
                     EFNRHGARYPSYNTIVGSGENGCILHYTENECEMRDGDLVLIDAGCEYKGYAGDITRT
                     FPVNGKFTQAQREIYDIVLESLETSLRLYRPGTSILEVTGEVVRIMVSGLVKLGILKG
                     DVDELIAQNAHRPFFMHGLSHWLGLDVHDVGVYGQDRSRILEPGMVLTVEPGLYIAPD
                     AEVPEQYRGIGIRIEDDIVITETGNENLTASVVKKPEEIEALMVAARKQ"
gene                 complement(9257. .9841)
                     /gene="ygfB"
CDS                  complement(9257. .9841)
                     /gene="ygfB"
                     /note="ORF_f194"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA69077.1"
                     /db_xref="PID:g882439"
                     /db_xref="GI:882439"
                     /translation="MLMSIQNEMPGYNEMNQYLNQQGTGLTPAEMHGLISGMICGGND
                     DSSWLPLLHDLTNEGMAFGHELAQALRKMHSATSDALQDDGFLFQLYLPDGDDVSVFD
                     RADALAGWVNHFLLGLGVTQPKLDKVTGETGEAIDDLRNIAQLGYDEDEDQEELEMSL
                     EEIIEYVRVAALLCHDTFTHPQPTAPEVQKPTLH"
CDS                  10003. .10332
                     /note="ORF_o109"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA69078.1"
                     /db_xref="PID:g882440"
                     /db_xref="GI:882440"
                     /translation="MSAQPVDIQIFGRSLRVNCPPDQRDALNQAADDLNQRLQDLKER
                     TRVTNTEQLVFIAALNISYELAQEKAKTRDYAASMEQRIRMLQQTIEQALLEQGRITE
                     KTNQNFE"
misc_feature         10266. .11219

  Query Match           15.6%;  Score 30.2;  DB 1;  Length 141744;
  Best Local Similarity  53.9%;  Pred. No. 21;
  Matches   62;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy      10 AGTGTAAGGACTCAGCCCCCCGTCATAATGCCAGGTATGCAGACCGAGATCAAAACGCAG 69
           || | | ||| | |||  |  |  | |||         || |   | ||| || || |
Db   65339 AGCGAACGGATTGAGCAACAGGAGACAATATTGCAGCAGCTGGTGGCGATGAACACCCGA 65280

Qy      70 GCCACGAATCTTGCGGCGAATCTTTCCGCAGTCAGAGAAAGTGCCACAGCGACGC 124
             | |||    ||| || || ||   ||||   || ||  ||||||| | ||||
Db   65279 CTGAAGAACGCCGCGCCGGATATTATTGCAGCGCGAAAATCTGCCACAACAACGC 65225


RESULT 11
AE000379/c
LOCUS       AE000379    10170 bp    DNA             BCT       12-NOV-1998
DEFINITION  Escherichia coli K-12 MG1655 section 269 of 400 of the complete
```

```
  Query Match             15.7%;  Score 30.4;  DB 5;  Length 4060;
  Best Local Similarity   51.5%;  Pred. No. 14;
  Matches   70;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy      55 GAGATCAAAACGCAGGCCACGAATCTTGCGGCGAATCTTTCCGCAGTCAGAGAAAGTGCC 114
           || |   |  | ||||    ||| || || | |||||      ||  |||||| |||| |
Db    3459 GAAACACAGCCCCAGGTGGAGAAACTGGCCGGGAATCAAGAGTCACCCAGAGACAGTGAC 3400

Qy     115 ACAGCGACGCTGTCAGGGGAAATTAAAGGCCCGCAACTGGAAGATTTTCCCGCGCTGATC 174
             | |  | ||||      |      |||    || |   || | |   | |  |||  |
Db    3399 CAACCATCCCTGTTTTCCTAGGACTGAGGGTTTCAGCGCTAAAACTAGGCTGTCCTGGGC 3340

Qy     175 AAACAGGCGAGTCTGG 190
           ||||||   |  ||||
Db    3339 AAACAGCATAAGCTGG 3324


RESULT  9
S51329/c
LOCUS       S51329       4067 bp    DNA             PRI       08-MAR-1993
DEFINITION  AAVS1=adeno-associated virus integration site [human, Genomic, 4067
            nt].
ACCESSION   S51329
NID         g261958
VERSION     S51329.1  GI:261958
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4067)
  AUTHORS   Kotin,R.M., Linden,R.M. and Berns,K.I.
  TITLE     Characterization of a preferred site on human chromosome 19q for
            integration of adeno-associated virus DNA by non-homologous
            recombination
  JOURNAL   EMBO J. 11 (13), 5071-5078 (1992)
  MEDLINE   93099882
  REMARK    GenBank staff at the National Library of Medicine created this
            entry [NCBI gibbsq 120789] from the original journal article.
            This sequence comes from Fig. 1A.
            Map location: 19q13.3-qter.
FEATURES             Location/Qualifiers
     source          1. .4067
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     gene            1. .4067
                     /partial
                     /gene="AAVS1"
                     /note="adeno-associated virus integration site"
BASE COUNT      542 a   1382 c   1179 g    962 t      2 others
ORIGIN


  Query Match             15.7%;  Score 30.4;  DB 10;  Length 4067;
  Best Local Similarity   51.5%;  Pred. No. 14;
  Matches   70;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy      55 GAGATCAAAACGCAGGCCACGAATCTTGCGGCGAATCTTTCCGCAGTCAGAGAAAGTGCC 114
           || |   |  | ||||    ||| || || | |||||      ||  |||||| |||| |
Db    3466 GAAACACAGCCCCAGGTGGAGAAACTGGCCGGGAATCAAGAGTCACCCAGAGACAGTGAC 3407

Qy     115 ACAGCGACGCTGTCAGGGGAAATTAAAGGCCCGCAACTGGAAGATTTTCCCGCGCTGATC 174
             | |  | ||||      |      |||    || |   || | |   | |  |||  |
Db    3406 CAACCATCCCTGTTTTCCTAGGACTGAGGGTTTCAGCGCTAAAACTAGGCTGTCCTGGGC 3347

Qy     175 AAACAGGCGAGTCTGG 190
           ||||||   |  ||||
Db    3346 AAACAGCATAAGCTGG 3331


RESULT  10
ECU28377/c
LOCUS       ECU28377   141744 bp    DNA             BCT       01-JUL-1995
DEFINITION  Escherichia coli K-12 genome; approximately 65 to 68 minutes.
ACCESSION   U28377
NID         g882431
VERSION     U28377.1  GI:882431
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 87107)
  AUTHORS   Plunkett,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JUN-1995) Guy Plunkett, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT     This sequence was determined as part of the E. coli Genome Project
            (Frederick R. Blattner, director) at the University of
            Wisconsin-Madison. Supported by award HG00301 from the NIH Human
            Genome Project. The entire sequence was independently determined
            from E. coli MG1655; overlaps with other sequence determinations
            are annotated. This entry should be considered somewhat
            provisional; it will be updated and merged with others at a later
            date.
FEATURES             Location/Qualifiers
     source          1. .141744
                     /organism="Escherichia coli"
                     /strain="K-12"
                     /sub_strain="MG1655"
                     /db_xref="taxon:562"
                     /map="approximately 65 to 68 minutes"
                     /note="This sequence comprises the following lambda
                     clones: DD656(EC18-164), DD658(EC17-126), DD660(EC30MM49),
                     DD663(EC19-62), DD665(EC27-689d), DD667(EC27-357),
                     DD670(EC30-45), DD674(EC27-1001), DD676(EC27-1160),
                     DD678(EC30K505-2), DD680(EC30K506-2), and DD682(EC17-140);
                     M13Janus vector was used for subcloning"
     CDS             complement(<1. .292)
                     /note="ORF_f97p"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA69070.1"
                     /db_xref="PID:g882432"
                     /db_xref="GI:882432"
                     /translation="MAIALVTGGSRGIGRATALLLAQEGYTVAVNYQQNLHAAQEVMN
                     LITQAGGKAFVLQADISDENQVVAMFTAIDQHDEPLAALVNNAGILFTQCTVE"
     misc_feature    complement(1. .5335)
                     /note="similar to GenBank Accession Number X73958
                     (ECGCVOP)"
     misc_feature    complement(70. .324)
                     /note="similar to GenBank Accession Number D21144
                     (ECOPYU18)"
     misc_feature    complement(318. .3634)
                     /note="similar to GenBank Accession Number L20872
                     (ECOGCVHP); bases 3318 to 3438 differ from the sequence
                     presented here, as well as from X73958 and D21144"
     gene            complement(559. .3432)
                     /gene="gcvP"
     CDS             complement(559. .3432)
                     /gene="gcvP"
                     /note="ORF_f957"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA69071.1"
                     /db_xref="PID:g882433"
                     /db_xref="GI:882433"
                     /translation="MTQTLSQLENSGAFIERHIGPDAAQQQEMLNAVGAQSLNALTGQ
                     IVPKDIQLATPPQVGAPATEYAALAELKAIASRNKRFTSYIGMGYTAVQLPPVILRNM
                     LENPGWYTAYTPYQPEVSQGRLEALLNFQQVTLDLTGLDMASASLLDEATAAAEAMAM
                     AKRVXKLKNANRFFVASDVHPQTLDVVRTRAETFGFEVIVDDAQKVLDHQDVFGVLLQ
                     QVGTTGEIHDYTALISELKSRKIVVSVAADIMALVLLTAPGKQGADIVFGSAQRFGVP
                     MGYGGPHAAFFAAKDEYKRSMPGRIIGVSKDAAGNTALRMAMQTREQHIRREKANSNI
                     CTSQVLLANIASLYAVYHGPVGLKRIANRIHRLTDILAAGLQQKGLKLRHAHYFDTLC
```

```
Db    1010 TCGAGACCGGCAGCCGCGCGAGCCTCCTGCTCTGCGGCGACAGCAGCCTGAGCAT 1064


RESULT   7
AF139177
LOCUS       AF139177     6062 bp    DNA             BCT       01-JUN-1999
DEFINITION  Streptomyces lividans aerial mycelium formation gene cluster; RamS
            (ramS), putative ATP binding membrane translocator (ramA), putative
            ATP binding membrane translocator (ramB), and putative two
            component response regulator (ramR) genes, complete cds.
ACCESSION   AF139177
NID         g4928925
VERSION     AF139177.1  GI:4928925
KEYWORDS    .
SOURCE      Streptomyces lividans.
  ORGANISM  Streptomyces lividans
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 6062)
  AUTHORS   Keijser,B.J.F., Canters,G.W. and Vijgenboom,E.
  TITLE     The role of the ram cluster in the copper-dependent differentiation
            of Streptomyces lividans 1326
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 6062)
  AUTHORS   Keijser,B.J.F., Canters,G.W. and Vijgenboom,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-1999) Leiden Institute of Chemistry, Leiden
            University, Einsteinweg 55, P.O.Box 9502, Leiden 2300 RA, The
            Netherlands
FEATURES             Location/Qualifiers
     source          1. .6062
                     /organism="Streptomyces lividans"
                     /strain="1326"
                     /db_xref="taxon:1916"
     gene            952. .1080
                     /gene="ramS"
     CDS             952. .1080
                     /gene="ramS"
                     /function="unknown"
                     /codon_start=1
                     /transl_table=11
                     /product="RamS"
                     /protein_id="AAD33774.1"
                     /db_xref="PID:g4928926"
                     /db_xref="GI:4928926"
                     /translation="MNLFDLQSMETPKEEAMGDVETGSRASLLLCGDSSLSITTCN"
     gene            1197. .3044
                     /gene="ramA"
     CDS             1197. .3044
                     /gene="ramA"
                     /note="RamA; ABC transporter family"
                     /codon_start=1
                     /transl_table=11
                     /product="putative ATP binding membrane translocator"
                     /protein_id="AAD33775.1"
                     /db_xref="PID:g4928927"
                     /db_xref="GI:4928927"
                     /translation="MRKSRPVPPRAPGRALEVLVLLCSVAAAVAAVAQPLALGRTLDL
                     LLRDGDAGWWLPLSAALLLGELLLDSATSLFTGRCNATWTASVRTRALRGLLRTAPEH
                     ARPYPPGDIGTRLTLNAADAGGAPAARAALAASLITPLGALVALALVDVWVALCVLTG
                     LPALALLLRSFARDTGATVAAYQRTQSLIASRLLEALEGADTIGAAGTGERERARVLA
                     PLAKLAAQGRHMWALHGRALGRSGVLVPLLTLAATAVGGLRLAAGELSVGDLLAVGRY
                     AQLTAGVGAAASLLGAIVRAREARRRTRELERMPATAYGTRRLPPNGPGELRLCGVRV
                     LRGGREVLRADGVRVPGGSTVAVVGRSGAGKSVLAAVAGRLIDPDEGHVLLDGVRLDR
                     LTHEALRTEVAYAFERPVLGEGTIAEAVADGARRSSRERVRQAARAAGADGFVRRLPH
                     GYDTPLPRAPLSGGEHQRLGLARAFAHAGRLLVLDDATSSLDTATEHEVDLALRRSVR
                     PGTRLVVAHRPSVADRADLVLWLEDGQVRAVGTHRELWHTAGYREVFGAGAGAGAGAG
                     AGAGAGADAGAGADAGPGSSPGPDSGAATAVGGSGPGPVRRPEPEEARP"
     gene            3041. .4855
                     /gene="ramB"
     CDS             3041. .4855
                     /gene="ramB"
                     /note="RamB; ABC transporter"
                     /codon_start=1
                     /transl_table=11
                     /product="putative ATP binding membrane translocator"
                     /protein_id="AAD33776.1"
                     /db_xref="PID:g4928928"
                     /db_xref="GI:4928928"
                     /translation="MTSTEAGALRRLAPPARRFLAHRKGVLIRLALWSLAESGQAFLV
                     GHAVARSVDEGFLAGDPRRGLLWLGVALVAVLSGARVVRGVFAQLAGVTEPLRDGLVR
                     HAVDRSMARAAPGGPGGTDRAAVSRLTNQVEIARDSFAGLVLTLRSFVFTAAGALLGL
                     LSLHPALLVVVLPPLAAGLALFLVTLRPMAAAQRRALAADEALGEHAASARAALRDLA
                     ACGTGPGAERHGADLVADAAAAARTLAGWAAVRTAALGVAGHLPVLALLVAVEWLRGH
                     GVSVGALLGAFTYLVQSLLPALHTLMTALGAAGSRLLVVLDRILGPEPEPEPEPEPEL
                     GSGLEPEPEPASEPESGPSTASAPAAAFAVHTAAAPAVELRSVTLSYGVRAEPVLDAL
                     DLRVAPGEHLAVVGPSGIGKSTLTRLVAGTLAPSRGEVRVAGRVVTGRPAAELAALRV
                     LVPQDAYVFSGTVGDNLAYLRTDPCPAELDAAVEAFGLAPLVERLGGLDATVRPAELS
                     PGERQLVALVRAYLSPAPLLLLDEATCHLDPASEARAEKALAGRSGTLVVVAHRLSSA
                     VRADRTLVLDGIRAQSGTHAELLGRSPLYRDLTGHWNS"
     gene            complement(4860. .5459)
                     /gene="ramR"
     CDS             complement(4860. .5459)
                     /gene="ramR"
                     /note="RamR"
                     /codon_start=1
                     /transl_table=11
                     /product="putative two component response regulator"
                     /protein_id="AAD33777.1"
                     /db_xref="PID:g4928929"
                     /db_xref="GI:4928929"
                     /translation="MVRIAVVHDEKLLRSALVQLLRSDDTLDVSSHCLDADGPELSAA
                     LPADVCVVDGECLTGPQDAGAGRLRARYGDRLVVLATAKRPGVLRRAFDGGALGLVDK
                     NAPAHRLITAVHTVARGERFLDETLTVALLKGAEMPLTTRELGVLTLASQGAPIAEIA
                     ARLHLSRGTVRNYMATAVRKVGARNRVDAIRIVQSAGWT"
BASE COUNT      681 a    2289 c    2298 g     794 t
ORIGIN


  Query Match             16.0%;  Score 31;  DB 2;  Length 6062;
  Best Local Similarity   48.6%;  Pred. No. 9.3;
  Matches   85;  Conservative    0;  Mismatches   90;  Indels    0;  Gaps    0;

Qy      18 GACTCAGCCCCCCGTCATAATGCCAGGTATGCAGACCGAGATCAAAACGCAGGCCACGAA 77
           ||||| |||||                |  ||  | ||| |||     |  |||| | |  | |
Db     890 GACTCGGCCCCACCAGGAGCCGTGAGAAAAGCACACCCCTTGCCGAACGAAAGGAATGCG 949

Qy      78 TCTTGCGGCGAATCTTTCCGCAGTCAGAGAAAGTGCCACAGCGACGCTGTCAGGGGAAAT 137
           || ||   |   ||   | ||||||   | |    ||  ||  |        |||| |
Db     950 TCATGAACCTGTTCGACCTGCAGTCGATGGAGACCCCCAAGGAAGAGGCCATGGGGGACG 1009

Qy     138 TAAAGGCCCGCAACTGGAAGATTTTCCCGCGCTGATCAAACAGGCGAGTCTGGAT 192
           |  || || ||| | |   ||   ||| || |||     ||||  |  |  | ||
Db    1010 TCGAGACCGGCAGCCGCGCGAGCCTCCTGCTCTGCGGCGACAGCAGCCTGAGCAT 1064


RESULT   8
I30049/c
LOCUS       I30049       4060 bp    DNA             PAT       13-DEC-1996
DEFINITION  Sequence 1 from patent US 5580703.
ACCESSION   I30049
NID         g1820840
VERSION     I30049.1  GI:1820840
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 4060)
  AUTHORS   Kotin,R.M., Berns,K.I. and Linden,R.M.
  TITLE     Human adeno-associated virus integration site DNA and uses thereof
  JOURNAL   Patent: US 5580703-A 1 03-DEC-1996;
FEATURES             Location/Qualifiers
     source          1. .4060
                     /organism="unknown"
BASE COUNT      542 a    1379 c    1176 g     961 t      2 others
ORIGIN
```

```
                     /gene="SC5A7.05"
     CDS             4351. .5016
                     /gene="SC5A7.05"
                     /note="SC5A7.05, unknown, len: 221 aa; highly similar to
                     the C-terminus of many ribA genes e.g. GCH2_BACSU GTP
                     cyclohydrolase II (EC 3.5.4.25) (398 aa), fasta sores;
                     opt: 656 z-score: 960.9 E(): 0, 53.1% identity in 196 aa
                     overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein SC5A7.05"
                     /protein_id="CAA19934.1"
                     /db_xref="PID:e1311947"
                     /db_xref="PID:g3334786"
                     /db_xref="GI:3334786"
                     /db_xref="SPTREMBL:O88011"
                     /translation="MTEKIGVLGKKTTQRTDVERIVVTPLPTVYGKFRAFGYFDHERG
```

```
  Query Match             16.0%;  Score 31;  DB 1;  Length 40337;
  Best Local Similarity   48.6%;  Pred. No. 11;
  Matches   85;  Conservative    0;  Mismatches   90;  Indels    0;  Gaps    0;

Qy      18 GACTCAGCCCCCCGTCATAATGCCAGGTATGCAGACCGAGATCAAAACGCAGGCCACGAA 77
           ||||| ||||| | |       | ||  | ||| |||     |  |||| | |  | |
Db   36247 GACTCGGCCCCACCAGGAGCCGTGAGAAAAGCACACCCCTTGCCGAACGAAAGGAATGCG 36306

Qy      78 TCTTGCGGCGAATCTTTCCGCAGTCAGAGAAAGTGCCACAGCGACGCTGTCAGGGGAAAT 137
           || ||   |   ||   | ||||||   | |    ||  ||  |        |||| |
Db   36307 TCATGAACCTGTTCGACCTGCAGTCGATGGAGACCCCCAAGGAAGAGGCCATGGGGGACG 36366

Qy     138 TAAAGGCCCGCAACTGGAAGATTTTCCCGCGCTGATCAAACAGGCGAGTCTGGAT 192
           |  || || ||| | |   ||   ||| || |||     ||||  |  |  | ||
Db   36367 TCGAGACCGGCAGCCGCGCGAGCCTCCTGCTCTGCGGCGACAGCAGCCTGAGCAT 36421
```

```
RESULT   6
SCU03771
LOCUS       SCU03771     6630 bp    DNA             BCT       13-SEP-1994
DEFINITION  Streptomyces coelicolor A3(2) aerial mycelium formation  gene
            cluster (ramA, ramB, and ramR) genes, complete cds.
ACCESSION   U03771
NID         g432991
VERSION     U03771.1  GI:432991
KEYWORDS    .
SOURCE      Streptomyces coelicolor.
  ORGANISM  Streptomyces coelicolor
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 6630)
  AUTHORS   Ma,H. and Kendall,K.
  TITLE     Cloning and analysis of a gene cluster from Streptomyces coelicolor
            that causes accelerated aerial mycelium formation in Streptomyces
            lividans
  JOURNAL   J. Bacteriol. 176 (12), 3800-3811 (1994)
  MEDLINE   94266734
REFERENCE   2  (bases 1 to 6630)
  AUTHORS   Kendall,K.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-NOV-1993) Kendall K.J., Tulane University, Cell and
            Molecular Biology, 2000 Percival Stern, New Orleans, LA 70118, USA
FEATURES             Location/Qualifiers
     source          1. .6630
                     /organism="Streptomyces coelicolor"
                     /strain="A3(2)"
                     /isolate="M130"
                     /db_xref="taxon:1902"
                     /clone="pS3"
     gene            1116. .3026
                     /gene="ramA"
     CDS             1116. .3026
                     /gene="ramA"
                     /codon_start=1
                     /transl_table=11
                     /product="potential ATP-binding membrane transport
                     protein"
                     /protein_id="AAA21388.1"
                     /db_xref="PID:g432992"
                     /db_xref="GI:432992"
                     /translation="MSPAAGAPGSGPGRRRLPALDVRRFSPVRKSRPVPPRAPGRALE
                     VLVLLCSVAAAVAAVAQPLALGRTLDLLLLRDGDAGWWLPLSAALLLGELLLDSATSLF
                     TGRCNATWTASVRTRALRGLLRTAPEHARPYPPGDIGTRLTLNAADAGGAPAARAALA
                     ASLITPLGALVALALVDVWVALCVLTGLPALALLLRSFARDTGATVAAYQRTQSLIAS
                     RLLEALEGADTIGAAGTGERERARVLAPLAELAAQGRHMWALHGRALGRSGVLVPLLT
                     LAATAVGGLRLAAGELSVGDLLAVGRYAQLTAGVGAAASLLGAIVRAREARRRTRELE
                     RMTATVYGTRRLPPNGPGELRLCGVRVLRGGREVLRAGGVRVPGGSTVAVVGRSGAGK
                     SVLAAVAGRLIDPDEGYVLLDGVRLDRLTHEALRTEVAYAFERPVLGEGTIAEAVADG
                     ARRSSRERVRQAARAAGADGFVRRLPHGYDTPLPRAPLSGGEHQRLGLARAFAHAGRL
                     LVLDDATSSLDTATEHEVDLALRRSVRPGTRLVVAHRPSVADRADLVLWLEDGQVRAV
                     GTHRELWHTAGYREVFGAGAGAGAGAGAGADAGAGADAGPGPDSGAATAVGGSGPG
                     PVRRPEPEEARP"
     gene            3023. .4849
                     /gene="ramB"
     CDS             3023. .4849
                     /gene="ramB"
                     /note="GTG start codon"
                     /codon_start=1
                     /transl_table=11
                     /product="potential ATP-binding membrane protein"
                     /protein_id="AAA21389.1"
                     /db_xref="PID:g432993"
                     /db_xref="GI:432993"
                     /translation="MTSTEAGALRRLAPPARRFLAHRKGVLVRLALWSLAESGQAFLV
                     GHAVARSVDEGFLAGDPRRGLLWLGVALVAVLSGARVVRGVFAQLAGVTEPLRDGLVR
                     HAVDRSMARAAPGGPGGTDRAAVSRLTNQVEIARDSFAGLVLTLRSFVFTAAGALLGL
                     LSLHPALLVVVLPPLAAGLALFLVTLRPMAAAQRRALAADEALGEHAASARAALRDLT
                     ACGTGPGAERHGADLVADAAAAARTLAGWAAVRTAALGVAGHLPVLALLVAVEWLRGH
                     GVSVGALLGAFTYLVQSLLPALHTLMTALGAAGSRLLVVLDRILGPEPEPEPEPEPEP
                     EPELGSGLEPEPEPASEPESGPSTASASAAAFAVHTAAAPAVELRSVTLSYGVRAEPV
                     LDALDLRVAPGEHLAVVGPSGIGKSTLTRLVAGTLAPSRGEVRVAGRVVTGRPAAELA
                     ALRVLVPQDAYVFSGTVGDNLAYLRTDPSPAELDAAVEAFGLAPLVERLGGLDATVRP
                     AELSPGERQLVALVRAYLSPAPLLLLDEATCHLDPASEARAEKALAGRSGTLVVVAHR
                     LSSAVRADRTLVLDGIRAQSGTHAELLGRSPLYRDLTGHWNS"
     gene            complement(4854. .5462)
                     /gene="ramR"
     CDS             complement(4854. .5462)
                     /gene="ramR"
                     /note="GTG start codon"
                     /codon_start=1
                     /transl_table=11
                     /product="potential response regulator for aerial mycelium
                     production"
                     /protein_id="AAA21390.1"
                     /db_xref="PID:g432994"
                     /db_xref="GI:432994"
                     /translation="MGEMVRIAVVHDEKLLRSALVQLLRSDDTLDVSSHCLDADGPEL
                     SAALPADVCVVDGECLTGPEDAGAGRLRARYGDRLVVLATAKRPGVLRRAFDGGALGL
                     VDKNAPAHRLITAVHTVARGERFLDETLTVALLKGAEMPLTTRELGVLTLASQGAPIA
                     EIAARLHLSRGTVRNYMATAVRKVGARNRVDAIRIVQSAGWT"
BASE COUNT      763 a   2482 c   2514 g    871 t
ORIGIN
```

```
  Query Match             16.0%;  Score 31;  DB 1;  Length 6630;
  Best Local Similarity   48.6%;  Pred. No. 9.4;
  Matches   85;  Conservative    0;  Mismatches   90;  Indels    0;  Gaps    0;

Qy      18 GACTCAGCCCCCCGTCATAATGCCAGGTATGCAGACCGAGATCAAAACGCAGGCCACGAA 77
           ||||| ||||| |       |  ||  | ||| |||     |  |||| | |  | |
Db    890 GACTCGGCCCCACCAGGAGCCGTGAGAAAAGCACACCCCTTGCCGAACGAAAGGAATGCG 949

Qy      78 TCTTGCGGCGAATCTTTCCGCAGTCAGAGAAAGTGCCACAGCGACGCTGTCAGGGGAAAT 137
           || ||   |   ||   | ||||||   | |    ||  ||  |        |||| |
Db    950 TCATGAACCTGTTCGACCTGCAGTCGATGGAGACCCCCAAGGAAGAGGCCATGGGGGACG 1009

Qy     138 TAAAGGCCCGCAACTGGAAGATTTTCCCGCGCTGATCAAACAGGCGAGTCTGGAT 192
           |  || || ||| | |   ||   ||| || |||     ||||  |  |  | ||
```

(URL; http://www.sanger.ac.uk/Projects/S_coelicolor/) CDS are numbered using the following system eg SC7B7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).
The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS.
Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at http://www.nih.go.jp/jun/cgi-bin/frameplot.pl. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or (att)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.
IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid 5A7 lies between 4G2 and 6G3 on the AseI-A genomic restriction fragment.

FEATURES Location/Qualifiers

source 1. .40337
/organism="Streptomyces coelicolor"
/strain="A3(2)"
/db_xref="taxon:1902"
/clone="cosmid 5A7"

gene 1. .1108
/gene="SC5A7.01"

CDS <1. .1108
/gene="SC5A7.01"
/note="SC5A7.01, possible glycosyl transferase, partial CDS, len: >368 aa; similar to several hypothetical proteins e.g. YV12_MYCTU MTCY20G9.12 (480 aa), fasta sores; opt: 237 z-score: 252.3 E(): 8.1e-07, 29.9% identity in 278 aa overlap, and to WCAL_ECOLI putative colanic acid biosynthesis glycosyl transferase (406 aa), fasta sores; opt: 187 z-score: 194.8 E(): 0.0013, 27.8% identity in 237 aa overlap. Contains Pfam match to entry PF00534 Glycos_transf_1, Glycosyl transferases group 1, score 52.90, E-value 5.5e-14"
/codon_start=2
/transl_table=11
/product="putative glycosyl transferase"
/protein_id="CAA19930.1"
/db_xref="PID:e1311943"
/db_xref="PID:g3334782"
/db_xref="GI:3334782"
/db_xref="SPTREMBL:O88007"
/translation="IVPMSLRAVHFVMPGGVDDPAAPSGGNAYDRRVRLDLPGFGWRV
RGLPVPGDWPRPDDAARAELARVLRRLPDGAVVLLDGLVACGVPEVVVPEAERLRMAV
LVHLPLGDETGLDPAVAAGLDARERTVLRAVPAVVATSDWAVRRLVSHHGLDPGRVHV
AAPGADIAPLAPGTDGVSKLVCVAAVTPRKGQHRLVEALAAVSDLPWSCVCVGGLAQD
PEYVDRLRSLIARHGLQERLVLAGPRAGADLDATYAAADLMVLTSYAETYGMAVTEAL
ARGIPVLATDVGGLPEAVGRAPDGGVPGILVPPEDPGALAAELRGWFGEPDVRRRLKA
AARGRRAALNGWASTAQSLAGVLHRLPGEPRRAA"

misc_feature 659. .1018
/gene="SC5A7.01"
/note="Pfam match to entry PF00534 Glycos_transf_1, Glycosyl transferases group 1, score 52.90, E-value 5.5e-14"

RBS 1094. .1099
/gene="SC5A7.01"
/note="possible RBS upstream of SC5A7.02"

gene 1105. .2490
/gene="SC5A7.02"

CDS 1105. .2490
/gene="SC5A7.02"
/note="SC5A7.02, unknown, len: 461 aa; contains 6 degenerate repeats of P(A/T)ESAVPA at the N-terminus"
/codon_start=1
/transl_table=11
/product="hypothetical protein SC5A7.02"
/protein_id="CAA19931.1"
/db_xref="PID:e1311944"
/db_xref="PID:g3334783"
/db_xref="GI:3334783"
/db_xref="SPTREMBL:O88008"
/translation="MNTSRPAPPTERTVPSPAESAVPASTESAVLARTGGAVPAPAES
VGPSPAGGTGPVRLGPPDATGSRIAVVPVSGPARSDAGRPVERPASGLGGGPGEGRAL
PSADRAEGGAARPATQRPGGADERTEDRAEQPSAGPVAEGQGQGRFGAGRPSAGSVAD
GLEQSSGERATVRLRGAPVREHPAEPPRYAPEWLRLREPADAAARAHDLLDPLRIRLA
DLPGRADGLVVHDLGCGTGSMGRWLAPLLDGAQHWVLHDRDPYLLHFAAAGAPHTAAD
GSRVTVETRRGDLAHLTPDALNGAGLVTASALLDVLTTEEVGKLAAACTAAGCPALLT
LSVAGRVDLTPGHPLDTEITEAFNDHQRRTGMLGPGAVTVATEAFAGHGATVRAQPSD
WRLGPGEAGLTAQWLRGWVGAAVEQRPELRARADRYLEERLAACAAGELRVVVHHTDL
LALCRPTGGAP"

repeat_region 1129. .1272
/note="6x 24 bp repeats"

RBS 2476. .2480
/gene="SC5A7.02"
/note="possible RBS upstream of SC5A7.03"

gene 2487. .3689
/gene="SC5A7.03"

CDS 2487. .3689
/gene="SC5A7.03"
/note="SC5A7.03, probable integral membrane protein, len: 400 aa. Note 63 aa overlap with convergent downstream ORF"
/codon_start=1
/transl_table=11
/product="hypothetical protein SC5A7.03"
/protein_id="CAA19932.1"
/db_xref="PID:e1311945"
/db_xref="PID:g3334784"
/db_xref="GI:3334784"
/db_xref="SPTREMBL:O88009"
/translation="MSAPTLPTVSTVPAAVPPTPVRPHAPTGVAAPRRRALRAHLGTL
AGTAVLAVLLWRLGTGAFLDGLRRIDGPAVLAALGIGLVTTVFSAWRWALVARGLRIR
LPLAAAVADYYRALFLNAALPGGVLGDVHRAVRHGRSTGDLGRGVRAVVLERVAGQLA
LIGAGAGVLLTMPSPVRAEVRHLAPLVALALAGALAVVLALRWNRAPARRGRAVRTAL
GEARHGLLSRRNLPGIVLSSAIVLAGHLTMFVVAARVAGSAASVAVLTPLAVLALLAM
GLPLNVGGWGPREGVTAWAFGAAGLGAGTGLGVAVVYGVLSLVASLPGVVVLVGRRYA
ARYPRSPSEPSEASESSEPSDVDGVPAVSAVSSATYSPKESARLSSSSFPFSADPSEG
RPMTPESV"

gene complement(3494. .4315)
/gene="SC5A7.04c"

CDS complement(3494. .4315)
/gene="SC5A7.04c"
/note="SC5A7.04c, unknown, len: 273 aa; some similarity to hypothetical proteins and to Pseudomonas TR:Q52548 (EMBL:D45424) creatinine amidohydrolase (259 aa), fasta sores; opt: 193 z-score: 273.1 E(): 5.6e-08 27.3% identity in 238 aa overlap. Note 63 aa overlap with convergent downstream ORF"
/codon_start=1
/transl_table=11
/product="hypothetical protein SC5A7.04c"
/protein_id="CAA19933.1"
/db_xref="PID:e1311946"
/db_xref="PID:g3334785"
/db_xref="GI:3334785"
/db_xref="SPTREMBL:O88010"
/translation="MKGRAGMDVSLVPADTSDDVRERGSGLPRQVAVLPVGSFEQHGP
YLPLATDTLVACAVARAVADAYPVHLLPPVTIGCSHEHADWPGTVSISAVTLHAVVRD
IADSLRRSGVGALVVVNGHGGNYVLGNVVQESSARGERMALFPAAEDWETARRAGVV
TSLLTDMHAGEIETSILLHKHPELVRPGYEAADFVADDRRHLLSLGMSAYTDSGVIGR
PSLGSAEKGKELLESLADSFGEYVALLTAETAGTPSTSEGSDDSEASDGSDGERG"

RBS 4335. .4338
/note="possible RBS upstream of SC5A7.05"

gene 4351. .5016

```
*   198099   198129: gap of unknown length
*   198130   199023: contig of 894 bp in length
*   199024   199054: gap of unknown length
*   199055   199947: contig of 893 bp in length
*   199948   199978: gap of unknown length
*   199979   200869: contig of 891 bp in length
*   200870   200900: gap of unknown length
*   200901   201785: contig of 885 bp in length
*   201786   201816: gap of unknown length
*   201817   202698: contig of 882 bp in length
*   202699   202729: gap of unknown length
*   202730   203611: contig of 882 bp in length
*   203612   203642: gap of unknown length
*   203643   204517: contig of 875 bp in length
*   204518   204548: gap of unknown length
*   204549   205419: contig of 871 bp in length
*   205420   205450: gap of unknown length
*   205451   206310: contig of 860 bp in length
*   206311   206341: gap of unknown length
*   206342   207200: contig of 859 bp in length
*   207201   207231: gap of unknown length
*   207232   208045: contig of 814 bp in length
*   208046   208076: gap of unknown length
*   208077   208888: contig of 812 bp in length
*   208889   208919: gap of unknown length
*   208920   209730: contig of 811 bp in length
*   209731   209761: gap of unknown length
*   209762   210569: contig of 808 bp in length
*   210570   210600: gap of unknown length
*   210601   211266: contig of 666 bp in length
*   211267   211297: gap of unknown length
*   211298   212067: contig of 770 bp in length
*   212068   212098: gap of unknown length
*   212099   212848: contig of 750 bp in length
*   212849   212879: gap of unknown length
*   212880   213625: contig of 746 bp in length
*   213626   213655: gap of unknown length
*   213656   214399: contig of 744 bp in length
*   214400   214429: gap of unknown length
*   214430   215171: contig of 742 bp in length
*   215172   215201: gap of unknown length
*   215202   215942: contig of 741 bp in length
*   215943   215972: gap of unknown length
*   215973   216710: contig of 738 bp in length
*   216711   216740: gap of unknown length
*   216741   217478: contig of 738 bp in length
*   217479   217508: gap of unknown length
*   217509   218242: contig of 734 bp in length
*   218243   218272: gap of unknown length
*   218273   218994: contig of 722 bp in length
*   218995   219024: gap of unknown length
*   219025   219742: contig of 718 bp in length
*   219743   219772: gap of unknown length
*   219773   220475: contig of 703 bp in length
*   220476   220505: gap of unknown length
*   220506   221198: contig of 693 bp in length
*   221199   221228: gap of unknown length
*   221229   221915: contig of 687 bp in length
*   221916   221945: gap of unknown length
*   221946   222631: contig of 686 bp in length
*   222632   222661: gap of unknown length
*   222662   223344: contig of 683 bp in length
*   223345   223374: gap of unknown length
*   223375   224056: contig of 682 bp in length
*   224057   224086: gap of unknown length
*   224087   224768: contig of 682 bp in length
*   224769   224798: gap of unknown length
*   224799   225466: contig of 668 bp in length
*   225467   225496: gap of unknown length
*   225497   226164: contig of 668 bp in length
*   226165   226194: gap of unknown length
*   226195   226801: contig of 607 bp in length
*   226802   226831: gap of unknown length
*   226832   227071: contig of 240 bp in length
*   227072   227101: gap of unknown length
*   227102   227751: contig of 650 bp in length
*   227752   227781: gap of unknown length
*   227782   228430: contig of 649 bp in length
*   228431   228460: gap of unknown length
*   228461   229105: contig of 645 bp in length
*   229106   229135: gap of unknown length
*   229136   229761: contig of 626 bp in length
*   229762   229791: gap of unknown length
*   229792   230346: contig of 555 bp in length.
FEATURES             Location/Qualifiers
     source          1. .230346
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RPCI11-303E5"
                     /chromosome="12p13.3"
BASE COUNT    65171 a  47004 c  49385 g  66532 t   2254 others
ORIGIN


  Query Match           16.1%;  Score 31.2;  DB 35;  Length 230346;
  Best Local Similarity   54.3%;  Pred. No. 11;
  Matches   63;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy      21 TCAGCCCCCCGTCATAATGCCAGGTATGCAGACCGAGATCAAAACGCAGGCCACGAATCT 80
           |||||| |||  ||  || |||| | ||||  |   ||    || || | |  ||  | |
Db  19876 TCAGCCTCCCTCCACCATCCCAGCTCTGCACCCATTGAAACTAAAGCTGCCATCGCCTTT 19817

Qy      81 TGCGGCGAATCTTTCCGCAGTCAGAGAAAGTGCCACAGCGACGCTGTCAGGGGAAA 136
           | |  | | || | || ||      || |  ||| |  |  || | |   |||||
Db  19816 TCCTTCCACTCCTCCCTCATAACCCTAAGGCCCCAAATAGGAGCAGACCCAGGAAA 19761


RESULT   5
SC5A7
LOCUS       SC5A7       40337 bp    DNA             BCT       27-JUL-1998
DEFINITION  Streptomyces coelicolor cosmid 5A7.
ACCESSION   AL031107
NID         g3334781
VERSION     AL031107.1  GI:3334781
KEYWORDS    6-phosphogluconate dehydrogenase; ABC transporter;
            glucose-6-phosphate 1-dehydrogenase; glucose-6-phosphate isomerase;
            glucosidase; glycosyl transferase; iron-sulfur oxidoreductase; pgi;
            ramA; ramB; ramR; secreted glucosidase; tal; tkt; transaldolase;
            transketolase; two-component regulator; two-component sensor; zwf.
SOURCE      Streptomyces coelicolor.
  ORGANISM  Streptomyces coelicolor
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 40337)
  AUTHORS   Seeger,K.J. and Harris,D.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 40337)
  AUTHORS   Parkhill,J., Barrell,B.G. and Rajandream,M.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JUL-1998) Streptomyces coelicolor sequencing project,
            Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
            David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
            Colney, Norwich, Norfolk NR4 7UH, UK
REFERENCE   3  (bases 1 to 40337)
  AUTHORS   Redenbach,M., Kieser,H.M., Denapaite,D., Eichner,A., Cullum,J.,
            Kinashi,H. and Hopwood,D.A.
  TITLE     A set of ordered cosmids and a detailed genetic and physical map
            for the 8 Mb Streptomyces coelicolor A3(2) chromosome
  JOURNAL   Mol. Microbiol. 21 (1), 77-96 (1996)
  MEDLINE   97000351
COMMENT     Notes:
            Streptomyces coelicolor sequencing at The Sanger Centre is funded
            by the BBSRC.
            Details of S. coelicolor sequencing at the Sanger Centre are
            available on the World Wide Web.
```

```
                DSKLEMDNAKYQAWQAGFKAQEENMKTTLQTLTQKYSNANSLYDNLVKVLSSTISSSW
                KPPKASCKDNRRGY"
     gene       1133. .3187
                /gene="sipA"
     CDS        1133. .3187
                /gene="sipA"
                /note="homolog of Shigella spp. IpaA"
                /codon_start=1
                /transl_table=11
                /product="SipA"
                /protein_id="AAA86618.1"
                /db_xref="PID:g1172128"
                /db_xref="GI:1172128"
                /translation="MVTSVRTQPPVIMPGMQTEIKTQATNLAANLSAVRAKCHSDAVR
                EIKGPQLEDFPALIKQASLDALFKCGKDAEALKEVFTNSNNVAGKKAIMEFAGLFRSA
                LNATSDSPEAKTLLMKVGAEYAAQIIKDGLKEKSAFGPWLPETKKAEAKLENLEKQLL
                DIIKNNTGGELSKLSTNLVMQEVMPYIASCIEHNFGCTLDPLTRSNLTHLVDKAAAKA
                VEALDMCPQKLTQEQGTSVGREARHLEMQTLIPLLLRNVFAQIPADKLPDPKIPEPAA
                GPVPDGGKKAEPTGINININIDSSNHSVDNSKHINNAEPVDNGQRHIDNSNHDNSRKT
                IDNSRTFIDNSQRNGESHHSTNSSNVSHSHSRVDSTTHQTETAHSASTGAIDHGIAGK
                IDVTAHATAEAVTNASSESKDGKVVTSEKGTTGETTSFDEVDGVTSKSIIGKPVQATV
                HGVDDNKQQSQTAEIVNVKPLASQLAGVENVKTDTLQSDTTVITGNKAGTTDNDNSQT
                DKTGPFSGLKFKQNSFLSTVPSVTNMHSMHFDARETFLGVIRKALEPDTSTPFPVRRA
                FDGLRAEILPNDTIKSAALKAQCSDIDKHPELKAKMETLKEVITHHPQKEKLAEIALQ
                FAREAGLTRLKGETDYVLSNVLDGLIGDGSWRAGPAYESYLNKPGVDRVITTVDGLHM
                QR"
     gene       3206. .3454
                /gene="iacP"
     CDS        3206. .3454
                /gene="iacP"
                /note="IacP"
                /codon_start=1
                /transl_table=11
                /product="acyl carrier protein"
                /protein_id="AAA86619.1"
                /db_xref="PID:g1172129"
                /db_xref="GI:1172129"
                /translation="MNMDIEARVKKVITSCIAVDVDSINGQTHLVEDLYADSLDLIDI
                VFGLSEEFDISCNENDLPDMMTFADICRVVKKSLESRV"
BASE COUNT     1048 a    759 c    879 g    814 t
ORIGIN


  Query Match           88.1%;  Score 171;  DB 1;  Length 3500;
  Best Local Similarity  99.0%;  Pred. No. 1.4e-43;
  Matches  193;  Conservative    0;  Mismatches    0;  Indels    2;  Gaps    2;

Qy      1 ATGGTTACAAGTGTAAGGACTCAGCCCCCCGTCATAATGCCAGGTATGCAGACCGAGATC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1133 ATGGTTACAAGTGTAAGGACTCAGCCCCCCGTCATAATGCCAGGTATGCAGACCGAGATC 1192

Qy     61 AAAACGCAGGCCACGAATCTTGCGGCGAATCTTTCCGCAGTCAGAG-AAAGTGCCACAGC 119
          ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db   1193 AAAACGCAGGCCACGAATCTTGCGGCGAATCTTTCCGCAGTCAGAGCAAAGTGCCACAGC 1252

Qy    120 GACGCTGTCAGGGGAAATTAAAGGCCCGCAACTGGAAGATTTTCCCGCGCTGATCAAACA 179
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db   1253 GACGCTGTCA-GGGAAATTAAAGGCCCGCAACTGGAAGATTTTCCCGCGCTGATCAAACA 1311

Qy    180 GGCGAGTCTGGATGC 194
          |||||||||||||||
Db   1312 GGCGAGTCTGGATGC 1326


RESULT  4
AC005842/c
LOCUS       AC005842   230346 bp    DNA             HTG       08-APR-1999
DEFINITION  Homo sapiens chromosome 12p13.3 clone RPCI11-303E5, WORKING DRAFT
            SEQUENCE, 65 unordered pieces.
ACCESSION   AC005842
NID         g4572651
VERSION     AC005842.5  GI:4572651
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 230346)
  AUTHORS   Muzny,D., Arenson,A.D., Brundage,E., Carvelli,K., Chen,E., Di,W.,
            Ding,Y., Dugan,S., Durbin,J., Forcum,J., Ganesh,R., Garcia,C.,
            Goodman,M., Gorrell,J.H., Haywood,M., Jackson,L., Kampal,R.,
            Karpathy,S., Leal,B., Li,Y., Liu,W., Logan,O., Lu,J., Ly,T.,
            Martinez,C., Oswal,G., Perez,L., Rashid,N.D., Rowland,K.,
            Savage,L., Scherer,S.S., Shen,H., Timms,K.M., Todd,J., Vo,Q.,
            Worley,K.C., Yu,W., Chinault,C., Nelson,D. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 230346)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-1998) Molecular and Human Genetics, Baylor
            College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Apr 7, 1999 this sequence version replaced gi:4454406.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 65 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *      1   153121: contig of 153121 bp in length
            *  153122   153152: gap of unknown length
            *  153153   168221: contig of 15069 bp in length
            *  168222   168252: gap of unknown length
            *  168253   176376: contig of 8124 bp in length
            *  176377   176407: gap of unknown length
            *  176408   177728: contig of 1321 bp in length
            *  177729   177759: gap of unknown length
            *  177760   179096: contig of 1337 bp in length
            *  179097   179127: gap of unknown length
            *  179128   180365: contig of 1238 bp in length
            *  180366   180396: gap of unknown length
            *  180397   181545: contig of 1149 bp in length
            *  181546   181576: gap of unknown length
            *  181577   182704: contig of 1128 bp in length
            *  182705   182735: gap of unknown length
            *  182736   183858: contig of 1123 bp in length
            *  183859   183889: gap of unknown length
            *  183890   184964: contig of 1075 bp in length
            *  184965   184995: gap of unknown length
            *  184996   186066: contig of 1071 bp in length
            *  186067   186097: gap of unknown length
            *  186098   187154: contig of 1057 bp in length
            *  187155   187185: gap of unknown length
            *  187186   188210: contig of 1025 bp in length
            *  188211   188241: gap of unknown length
            *  188242   189245: contig of 1004 bp in length
            *  189246   189276: gap of unknown length
            *  189277   190272: contig of 996 bp in length
            *  190273   190303: gap of unknown length
            *  190304   191290: contig of 987 bp in length
            *  191291   191321: gap of unknown length
            *  191322   192300: contig of 979 bp in length
            *  192301   192331: gap of unknown length
            *  192332   193309: contig of 978 bp in length
            *  193310   193340: gap of unknown length
            *  193341   194295: contig of 955 bp in length
            *  194296   194326: gap of unknown length
            *  194327   195268: contig of 942 bp in length
            *  195269   195299: gap of unknown length
            *  195300   196228: contig of 929 bp in length
            *  196229   196259: gap of unknown length
            *  196260   197172: contig of 913 bp in length
            *  197173   197203: gap of unknown length
            *  197204   198098: contig of 895 bp in length
```

```
                YEFYNQGRLDEAETFFRFLCIYDFYNPDYTMGLAAVCQLKKQFQKACDLYAVAFTLLK
                NDYRPVFFTGQCQLLMRKAAKARQCFELVNERTEDESLRAKALVYLEALKTAETEQHS
                EQEKE"
gene            543. .2324
                /gene="sipB"
CDS             543. .2324
                /gene="sipB"
                /codon_start=1
                /transl_table=11
                /protein_id="CAA57988.1"
                /db_xref="PID:g1009003"
                /db_xref="GI:1009003"
                /db_xref="SPTREMBL:Q56134"
                /translation="MVNDASSISRSGYTQNPRLAEAAFEGVRKNTDFLKAADKAFKDV
                VATKAGDLKAGTKSGESAINTVGLKPPTDAAREKLSSEGQLTLLLGKLMTLLGDVSLS
                QLESRLAVWQAMIESQKEMGIQVSKEFQTALGEAQEATDLYEASIKKTDTAKSVYDAA
                AKKLTQAQNKLQSLDPADPGYAQAEAAVEQAGKEATEAKEALDKATDATVKAGTDAKA
                KAEKADNILTKFQGTANAASQNQVSQGEQDNLSNVARLTMLMAMFIEIVGKNTEESLQ
                NDLALFNALQEGRQAEMEKKSAEFQEETRKAEETNRIMGCIGKVLGALLTIVSVVAAV
                FTGGASLALAAVGLAVMVADEIVKAATGVSFIQQALNPIMEHVLKPLMELIGKAITKA
                LEGLGVDKKTAEMAGSIVGAIVAAIAMVAVIVVVAVVGKGAAAKLGNALSKMMGETIK
                KLVPNVLKQLAQNGSKLFTQGMQRITSGLGNVGSKMGLQTNALSKELVGNTLNKVALG
                MEVTNTAAQSAGGVAEGVFIKNASEALADFMLARFAMDQIQQWLKQSVEIFGENQKVT
                AELQKAMSSAVQQNADASRFILRQSRA"
gene            2352. .3581
                /gene="sipC"
CDS             2352. .3581
                /gene="sipC"
                /codon_start=1
                /transl_table=11
                /protein_id="CAA57989.1"
                /db_xref="PID:g1009004"
                /db_xref="GI:1009004"
                /db_xref="SPTREMBL:Q56135"
                /translation="MLISNVGINPAAYLNNHSVENSSQTASQSVSAKDILNSIGISSS
                KVSDLGLSPTLSAPAPGVLTQTPGTITSFLKASIQNTDMNQDLNALANNVTTKANEVV
                QTQLREQQAEVGKFFDISGMSSSAVALLAAANTLMLTLNQADSKLSGKLSLVSFDAAK
                TTASSMMREGMNALSGSISQSALQLGITGVGAKLEYKGLQNERGALKHNAAKIDKLTT
                ESHSIKNVLNGQNSVKLGAEGVDSLKSLNMKKTGTDATKNLNDATLKSNAGTSATESL
                GIKNSNKQISPEHQAILSKRLESVESDIRLEQNTMDMTRIDARKMQMTGDLIMKNSVT
                VGGIAGASRQYAATQERSEQQISQVNNRVASTASDEARESSRKSTSLIQEMLKTMESI
                NQSKASALAAIAGNIRA"
gene            3651. .4670
                /gene="sipD"
CDS             3651. .4670
                /gene="sipD"
                /codon_start=1
                /transl_table=11
                /protein_id="CAA57990.1"
                /db_xref="PID:g1009005"
                /db_xref="GI:1009005"
                /db_xref="SPTREMBL:Q56136"
                /translation="MLNIQNYSASPHPGIVAERPQTPSASEHAEIAVVPSTTEHRGTD
                IISLSQAATKIQQAQQTLQSTPPISEENNDERTLARQQLTSSLNALAKSGVSLSAEQN
                ENLRSTFSARRRPYLALRLWPARTTISDAEIWDMVSQNISAIGDSYLGVYENVVAVYT
                DFYQAFSDILSKMGGWLSPGKDGNTIKLNVDSLKSEISSLINKYTQINKNTILFPSQT
                GSGMTTATKAEAEQWIKELNLPDSCLKASGSGYVVLVDTGPLSKMVSDLNGIGSGSAL
                ELDNAKYQAWQSGFKAQEENLKTTLQTLTQKYSNANSLYDNLVKVLSSTISSSLETAK
                SFLQG"
gene            4689. .5393
                /gene="sipA"
CDS             4689. .>5393
                /gene="sipA"
                /codon_start=1
                /transl_table=11
                /protein_id="CAA57991.1"
                /db_xref="PID:g1009006"
                /db_xref="GI:1009006"
                /db_xref="SPTREMBL:Q56137"
                /translation="MVTSVRTQPPVIMPGMQTEIKTQATNLAANLSAVRESATATLSG
                EIKGQQLEDFPALIKQASLDALFKCGKDAEALKEVFTNSNNVAGKKAIMEFAGLFRSA
                LNATSDSPEAKTLLMKVGAEYTAQIIKDGLKEKSAFGPWLPETKKAEAKLENLEKQLL
                DIIKNNTGGELSKLSTNLVMQEVMPYIASCIEHNFGCTLDPLTRSSLTQLVDKAAAKA
                VEALDMCHQKLTQEQGT"
BASE COUNT      1584 a   1156 c   1458 g   1195 t
ORIGIN


  Query Match             96.7%;  Score 187.6;  DB 1;  Length 5393;
  Best Local Similarity   97.9%;  Pred. No. 9.1e-49;
  Matches  190;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 ATGGTTACAAGTGTAAGGACTCAGCCCCCCGTCATAATGCCAGGTATGCAGACCGAGATC 60
           |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    4689 ATGGTTACAAGTGTAAGAACTCAGCCCCCCGTCATAATGCCAGGTATGCAGACCGAGATC 4748

Qy      61 AAAACGCAGGCCACGAATCTTGCGGCGAATCTTTCCGCAGTCAGAGAAAGTGCCACAGCG 120
           |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db    4749 AAAACGCAGGCCACGAATCTTGCGGCGAATCTTTCCGCGGTCAGAGAAAGTGCCACAGCA 4808

Qy     121 ACGCTGTCAGGGGAAATTAAAGGCCCGCAACTGGAAGATTTTCCCGCGCTGATCAAACAG 180
           |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db    4809 ACGCTGTCAGGGGAAATTAAAGGCCAGCAACTGGAAGATTTTCCCGCGCTGATCAAACAG 4868

Qy     181 GCGAGTCTGGATGC 194
           ||||||||||||||
Db    4869 GCGAGTCTGGATGC 4882


RESULT   3
STU40013
LOCUS       STU40013     3500 bp    DNA             BCT       02-FEB-1996
DEFINITION  Salmonella typhimurium sipD, sipA, and acyl carrier protein (iacP)
            genes, complete cds.
ACCESSION   U40013
NID         g1172126
VERSION     U40013.1  GI:1172126
KEYWORDS    .
SOURCE      Salmonella typhimurium strain=SL1344.
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Salmonella.
REFERENCE   1  (bases 1 to 3500)
  AUTHORS   Kaniga,K., Trollinger,D. and Galan,J.E.
  TITLE     Identification of two targets of the type III protein secretion
            system encoded by the inv and spa loci of Salmonella typhimurium
            that have homology to the Shigella IpaD and IpaA proteins
  JOURNAL   J. Bacteriol. 177 (24), 7078-7085 (1995)
  MEDLINE   96099289
REFERENCE   2  (bases 1 to 3500)
  AUTHORS   Galan,J.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-NOV-1995) Jorge E. Galan, Molecular Gentics and
            Microbiology, School of Medicine, SUNY Stony Brook, Stony Brook, NY
            11794-5222, USA
FEATURES             Location/Qualifiers
     source          1. .3500
                     /organism="Salmonella typhimurium"
                     /strain="SL1344"
                     /db_xref="taxon:602"
     gene            84. .1130
                     /gene="sipD"
     CDS             84. .1130
                     /gene="sipD"
                     /note="homolog of Shigella spp. IpaD"
                     /codon_start=1
                     /transl_table=11
                     /product="SipD"
                     /protein_id="AAA86617.1"
                     /db_xref="PID:g1172127"
                     /db_xref="GI:1172127"
                     /translation="MLNIQNYSASPHPGIVAERPQTPSASEHVETAVVPSTTEHRGTD
                     IISLSQAATKIHQAQQTLQSTPPISEENNDERTLARQQLTSSLNALAKSGVSLSAEQN
                     ENLRSAFSAPTSALFSASPMAQPRTTISDAEIWDMVSQNISAIGDSYLGVYENVVAVY
                     TDFYQAFSDILSKMGGWLLPGKDGNTVKLDVTSLKNDLNSLVNKYNQINSNTVLFPAQ
                     SGSGVKVATEAEARQWLSELNLPNSCLKSYGSGYVVTVDLTPLQKMVQDIDGLGAPGK
```

```
          Massachusetts General Hospital, Harvard Medical School, Fruit
          Street, Boston, MA 02114, USA
FEATURES             Location/Qualifiers
     source          1. .3441
                     /organism="Salmonella typhimurium"
                     /strain="SL1344"
                     /db_xref="taxon:602"
                     /map="59 min"
     gene            1. .870
                     /gene="sspB"
     CDS             <1. .870
                     /gene="sspB"
                     /note="secreted protein"
                     /codon_start=1
                     /transl_table=11
                     /function="invasion determinant"
                     /product="SspB"
                     /protein_id="AAC43546.1"
                     /db_xref="PID:g1117812"
                     /db_xref="GI:1117812"
                     /translation="RKAEETNRIMGCIGKVLGALLTIVSVVAAVFTGGASLALAAVGL
                     AVMVADEIVKAATGVSFIQQALNPIMEHVLKPLMELIGKAITKALEGLGVDKKTAEMA
                     GSIVGAIVAAIAMVAVIVVVAVVGKGAAAKLGNALSKMMGETIKKLVPNVLKQLAQNG
                     SKLFTQGMQRITSGLGNVGSKMGLQTNALSKELVGNTLNKVALGMEVTNTAAQSAGGV
                     AEGVFIKNASEALADFMLARFAMDQIQQWLKQSVEIFGENQKVTAELQKAMSSAVQQN
                     ADASRFILRQSRA"
     gene            898. .2127
                     /gene="sspC"
     CDS             898. .2127
                     /gene="sspC"
                     /note="secreted protein"
                     /codon_start=1
                     /transl_table=11
                     /function="invasion determinant"
                     /evidence=experimental
                     /product="SspC"
                     /protein_id="AAC43547.1"
                     /db_xref="PID:g1117813"
                     /db_xref="GI:1117813"
                     /translation="MLISNVGINPAAYLNNHSVENSSQTASQSVSAKDILNSIGISSS
                     KVSDLGLSPTLSAPAPGVLTQTPGTITSFLKASIQNTDMNQDLNALANNVTTKANEVV
                     QTQLREQQAEVGKFFDISGMSSSAVALLAAANTLMLTLNQADSKLSGKLSLVSFDAAK
                     TTASSMMREGMNALSGSISQSALQLGITGVGAKLEYKGLQNERGALKHNAAKIDKLTT
                     ESHSIKNVLNGQNSVKLGAEGVDSLKSLNMKKTGTDATKNLNDATLKSNAGTSATESL
                     GIKDSNKQISPEHQAILSKRLESVESDIRLEQNTMDMTRIDARKMQMTGDLIMKNSVT
                     VGGIAGASGQYAATQERSEQQISQVNNRVASTASDEARESSRKSTSLIQEMLKTMESI
                     NQSKASALAAIAGNIRA"
     gene            2198. .3229
                     /gene="sspD"
     CDS             2198. .3229
                     /gene="sspD"
                     /note="secreted protein"
                     /codon_start=1
                     /transl_table=11
                     /function="invasion determinant"
                     /evidence=experimental
                     /product="SspD"
                     /protein_id="AAC43548.1"
                     /db_xref="PID:g1117814"
                     /db_xref="GI:1117814"
                     /translation="MLNIQNYSASPHPGIVAERPQTPSASEHVETAVVPSTTEHRGTD
                     IISLSQAATKIHQAQQTLQSTPPISEENNDERTLARQQLTSSLNALAKSGVSLSAEQN
                     ENLRSAFSAPTSALFSASPMAQPRTTISDAEIWDMVSQNISAIGDSYLGVYENVVAVY
                     TDFYQAFSDILSKMGGWLLPGKDGNTVKLDVTSLKNDLNSLVNKYNQINSNTVLFPAQ
                     SGSGVKVATEAEARQWLSELNLPNSCLKSYGSGYVVTVDLTPLQKMVQDIDGLGAPGK
                     DSKLEMDNAKYQAWQSGFKAQEENMKTTLQTLTQKYSNANSLYDNLVKVLSSTISSSL
                     ETAKSFLQG"
     gene            3248. .3441
                     /gene="sspA"
     CDS             3248. .>3441
                     /gene="sspA"
                     /note="secreted protein"
                     /codon_start=1
                     /transl_table=11
                     /function="unknown"
                     /product="SspA"
                     /protein_id="AAC43549.1"
                     /db_xref="PID:g1117815"
                     /db_xref="GI:1117815"
                     /translation="MVTSVRTQPPVIMPGMQTEIKTQATNLAANLSAVRESATATLSG
                     EIKGPQLEDFPALIKQASLDA"
BASE COUNT      978 a    759 c    925 g    779 t
ORIGIN


  Query Match             100.0%;  Score 194;  DB 2;  Length 3441;
  Best Local Similarity   100.0%;  Pred. No. 8.7e-51;
  Matches  194;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ATGGTTACAAGTGTAAGGACTCAGCCCCCCGTCATAATGCCAGGTATGCAGACCGAGATC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3248 ATGGTTACAAGTGTAAGGACTCAGCCCCCCGTCATAATGCCAGGTATGCAGACCGAGATC 3307

Qy     61 AAAACGCAGGCCACGAATCTTGCGGCGAATCTTTCCGCAGTCAGAGAAAGTGCCACAGCG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3308 AAAACGCAGGCCACGAATCTTGCGGCGAATCTTTCCGCAGTCAGAGAAAGTGCCACAGCG 3367

Qy    121 ACGCTGTCAGGGGAAATTAAAGGCCCGCAACTGGAAGATTTTCCCGCGCTGATCAAACAG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3368 ACGCTGTCAGGGGAAATTAAAGGCCCGCAACTGGAAGATTTTCCCGCGCTGATCAAACAG 3427

Qy    181 GCGAGTCTGGATGC 194
          ||||||||||||||
Db   3428 GCGAGTCTGGATGC 3441


RESULT   2
STSIPGNS
LOCUS       STSIPGNS     5393 bp    DNA             BCT       04-OCT-1995
DEFINITION  S.typhi spaT, sipB, sipC, sipD and sipA genes.
ACCESSION   X82670
NID         g1009001
VERSION     X82670.1  GI:1009001
KEYWORDS    sipA gene; sipB gene; sipC gene; sipD gene; spaT gene.
SOURCE      Salmonella typhi.
  ORGANISM  Salmonella typhi
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Salmonella.
REFERENCE   1  (bases 1 to 5393)
  AUTHORS   Hermant,D., Menard,R., Arricau,N., Parsot,C. and Popoff,M.Y.
  TITLE     Functional conservation of the Salmonella and Shigella effectors of
            entry into epithelial cells
  JOURNAL   Mol. Microbiol. 17 (4), 781-789 (1995)
  MEDLINE   96111497
REFERENCE   2  (bases 1 to 5393)
  AUTHORS   Popoff,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-1994) M. Popoff, Institut Pasteur, Unite des
            Enterobacteries, 28 rue du Docteur Roux, 75724 Paris Cedex 15,
            FRANCE
FEATURES             Location/Qualifiers
     source          1. .5393
                     /organism="Salmonella typhi"
                     /strain="Ty2"
                     /db_xref="taxon:601"
     gene            43. .540
                     /gene="spaT"
     CDS             43. .540
                     /gene="spaT"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA57987.1"
                     /db_xref="PID:g1009002"
                     /db_xref="GI:1009002"
                     /db_xref="SPTREMBL:Q57260"
                     /translation="MDYQNNVSEERVAEMIWDAVSEGATLKDVHGIPQDMMDGLYAHA
```

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 1999, 18:01:01 ; Search time 2414.69 Seconds
(without alignments)
255.511 Million cell updates/sec

Title: US-09-068-804-4
Perfect score: 194
Sequence: 1 ATGGTTACAAGTGTAAGGAC.....AAACAGGCGAGTCTGGATGC 194

Scoring table: IDENTITY_NUC

Searched: 679419 seqs, 1590154680 residues

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_st:*
14: gb_sts:*
15: gb_sy:*
16: gb_un:*
17: gb_vi:*
18: em_fun:*
19: em_htg:*
20: em_hum1:*
21: em_hum2:*
22: em_in:*
23: em_om:*
24: em_or:*
25: em_ov:*
26: em_pat:*
27: em_ph:*
28: em_pl:*
29: em_ro:*
30: em_sts:*
31: em_sy:*
32: em_un:*
33: em_vi:*
34: gb_htg1:*
35: gb_htg2:*
36: gb_in1:*
37: gb_in2:*
38: em_ba1:*
39: em_ba2:*
40: em_hum3:*
41: em_hum4:*
42: gb_pr4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	194	100.0	3441	2	STU30491	U30491 Salmonella
2	187.6	96.7	5393	1	STSIPGNS	X82670 S.typhi spa
3	171	88.1	3500	1	STU40013	U40013 Salmonella
c 4	31.2	16.1	230346	35	AC005842	AC005842 Homo sapi
5	31	16.0	40337	1	SC5A7	AL031107 Streptomy
6	31	16.0	6630	1	SCU03771	U03771 Streptomyce
7	31	16.0	6062	2	AF139177	AF139177 Streptomy
c 8	30.4	15.7	4060	5	I30049	I30049 Sequence 1
c 9	30.4	15.7	4067	10	S51329	S51329 AAVS1=adeno
c 10	30.2	15.6	141744	1	ECU28377	U28377 Escherichia
c 11	30.2	15.6	10170	2	AE000379	AE000379 Escherich
12	30.2	15.6	110000	34	CEY53C10_2	Continuation (3 of
c 13	30.2	15.6	175591	35	AC007606	AC007606 Homo sapi
14	30.2	15.6	78153	36	CEY53C10A	AL033536 Caenorhab
c 15	29.8	15.4	206606	8	ATFCA1	Z97336 Arabidopsis
c 16	29.6	15.3	170427	34	AC006095	AC006095 Homo sapi
17	29.4	15.2	4080	2	AF068771	AF068771 Synechocy
18	29.2	15.1	120206	34	HS439F8	AL021392 Homo sapi
c 19	29	14.9	128765	11	HUAC002040	AC002040 Homo sapi
c 20	29	14.9	1768	36	PPZ80955	Z80955 P.pulvillus
c 21	28.8	14.8	1477	1	D84127	D84127 Clavibacter
22	28.8	14.8	35197	11	AC005782	AC005782 Homo sapi
c 23	28.6	14.7	476	12	HAMHMGC03	L00326 Syrian hams
c 24	28.4	14.6	130690	1	D90901	D90901 Synechocyst
25	28.4	14.6	338534	1	ECOUW93	U14003 Escherichia
26	28.4	14.6	1832	1	ECSMSG	X63155 E.coli sms
27	28.4	14.6	10589	2	AE000509	AE000509 Escherich
28	28.4	14.6	1432	2	ECU59449	U59449 Escherichia
29	28.4	14.6	3935	7	FONIAGNA	Z22549 F.oxysporum
c 30	28.4	14.6	139488	11	AC003043	AC003043 Homo sapi
31	28.4	14.6	144734	34	AC005052	AC005052 Homo sapi
c 32	28.4	14.6	189805	42	AC004169	AC004169 Homo sapi
33	28.2	14.5	3608	12	RNU18942	U18942 Rattus norv
34	28	14.4	537	4	ABAJ822	AJ222822 Acipenser
35	28	14.4	3505	5	AR001476	AR001476 Sequence
36	28	14.4	141984	8	AC007591	AC007591 Arabidops
c 37	28	14.4	3612	12	CGU09453	U09453 Cricetulus
38	28	14.4	5787	12	RATAMPASGB	M38061 Rattus norv
39	28	14.4	3488	12	RATGLUR2A	M85035 Rat glutama
40	28	14.4	3120	12	RNGLURK2	X54655 Rat GluR-K2
41	28	14.4	494	14	HUM4S1017	L14083 Human chrom
42	28	14.4	139675	34	HS564M11	AL035409 Homo sapi
c 43	27.8	14.3	215534	1	BSUB0021	Z99124 Bacillus su
44	27.8	14.3	65143	1	D83026	D83026 Bacillus su
c 45	27.8	14.3	277	1	UAU35704	U35704 Unidentifie

ALIGNMENTS

RESULT 1
STU30491
LOCUS STU30491 3441 bp DNA BCT 19-JAN-1996
DEFINITION Salmonella typhimurium SspA (sspA) and SspB (sspB) genes, partial cds, and SspC (sspC) and SspD (sspD) genes, complete cds.
ACCESSION U30491
NID g1117811
VERSION U30491.1 GI:1117811
KEYWORDS .
SOURCE Salmonella typhimurium.
ORGANISM Salmonella typhimurium
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Salmonella.
REFERENCE 1 (bases 1 to 3441)
AUTHORS Hueck,C.J., Hantman,M.J., Bajaj,V., Johnston,C., Lee,C.A. and Miller,S.I.
TITLE Salmonella typhimurium secreted invasion determinants are homologous to Shigella Ipa proteins
JOURNAL Mol. Microbiol. 18 (3), 479-490 (1995)
MEDLINE 96342378
REFERENCE 2 (bases 1 to 3441)
AUTHORS Miller,S.I.
TITLE Direct Submission
JOURNAL Submitted (28-JUN-1995) Samuel I. Miller, Infectious Disease Unit,

```
Db    157 GCACTAATCATAGAAAGCATGACATTGAGCCTTAAAAGGCTAAGAACGCACATGCAAATG 98

Qy    154 GAAG 157
Db     97 ATAG 94
```

RESULT 14
Q78090/c
ID Q78090 standard; cDNA; 280 BP.
AC Q78090;
DT 15-AUG-1995 (first entry)
DE Hepatitis C virus NS5B region.
KW Hepatitis C virus; HCV; primer; probe; detection; diagnosis;
KW classification; immunisation; prophylaxis; serotyping; ss.
OS Hepatitis C virus.
FH Key Location/Qualifiers
FT cds 2. .280
FT /*tag= a
FT /product= NS5B polypeptide.
FT mat_peptide 2. .277
FT /*tag= b
PN WO9425601-A.
PD 10-NOV-1994.
PF 27-APR-1994; E01323.
PR 27-APR-1993; EP-401099.
PR 05-AUG-1993; EP-402019.
PA (INNO-) INNOGENETICS NV SA.
PI Maertens G, Stuyver L;
DR WPI; 94-358277/44.
DR P-PSDB; R63352.
PT New polynucleotide sequences from hepatitis C virus - and related
PT vectors, polypeptide(s) and antibodies, useful for immunisation,
PT treatment, diagnosis and typing of HCV isolates
PS Claim 2; Page 199; 404pp; English.
CC Compositions comprising at least 5, and pref. 8 or more contiguous
CC nucleotides selected from an HCV type 3 genomic sequence, more
CC particularly (i) the region spanning positions 417-957 of the
CC Core/E1 region of HCV subtype 3a; (ii) the region spanning positions
CC 4664-4730 of the NS3 region of HCV type 3; (iii) the region spanning
CC positions 4892-5292 of the NS3/4 region of HCV type 3; (iv) the
CC region spanning positions 8023-8235 of the NS5 region of the BR36
CC subgroup of HCV subtype 3a; or (v) an HCV subtype 3c genomic
CC sequence, may be used as primers to amplify nucleic acid from an
CC isolate belonging to a specific genotype, or as a probe for specific
CC detection/classification of nucleic acid. Polypeptides encoded by
CC the nucleotides in such compositions may be used for immunisation
CC against HCV, for the detection of antibodies directed against HCV
CC and for serotyping. This sequence corresponds to the NS5B region
CC of HCV.
SQ Sequence 280 BP; 70 A; 69 C; 82 G; 59 T;

Query Match 13.5%; Score 26.2; DB 1; Length 280;
Best Local Similarity 49.6%; Pred. No. 7.8;
Matches 67; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

```
Qy     36 AATGCCAGGTATGCAGACCGAGATCAAAACGCAGGCCACGAATCTTGCGGCGAATCTTTC 95
Db    256 ACTCTCAGCAATCACGACCAAGTCATCTCCGCAGACAAGGAATGATGGATTTTTAATTCC 197

Qy     96 CGCAGTCAGAGAAAGTGCCACAGCGACGCTGTCAGGGGAAATTAAAGGCCCGCAACTGGA 155
Db    196 GGCAGCTTTGGTAGCTGCCTTTGCTTTGATGTAACAGGTGATTGTGTTCCCGAAGCTGGT 137

Qy    156 AGATTTTCCCGCGCT 170
Db    136 GGGTAGCACCCCACT 122
```

RESULT 15
T27985/c
ID T27985 standard; DNA; 340 BP.
AC T27985;
DT 14-MAR-1997 (first entry)
DE Hepatitis C virus type 10a isolate NE98 bases 7932-8271.
KW Hepatitis C virus; subtype; polymerase chain reaction; amplification;
KW PCR; primer; probe; antibody; infection; ss.
OS Hepatitis C virus.
FH Key Location/Qualifiers
FT cds 2. .340
FT /*tag= a
FT /note= "encodes amino acids 2645-2757 of HCV NS5B
FT protein"
PN WO9613590-A2.
PD 09-MAY-1996.
PF 23-OCT-1995; E04155.
PR 21-OCT-1994; EP-870166.
PR 28-JUN-1995; EP-870076.
PA (INNO-) INNOGENETICS NV.
PI Maertens G, Stuyver L;
DR WPI; 96-251460/25.
DR P-PSDB; R96574.
PT Hepatitis C virus poly:nucleic acid unique to unidentified sub:type
PT - used to develop probes and primers for new sub:types and vaccines
PT to prevent and treat infection
PS Claim 6; Fig 3; 150pp; English.
CC The sequences T27937-T27989 represent novel sequences isolated from
CC hepatitis C virus subtypes different from subtypes 1a-c, 2a-d, 3a-f,
CC 4a-j, 5a and 6a. They esp. from the novel subtypes 1d-f, 2e-i, 2k, 2l,
CC 3g, 4k-m, 7a-c or types 9, 10 or 11. The sequences corresp. to the 5'
CC untranslated region (UR), the Core/E1, NS4 or NS5B regions of the genome.
CC This sequence represents nucleotides 7932-8271 from the HCV type 10a
CC isolate NE98.
CC The new HCV types were isolated from patients with chronic HCV from the
CC Benelux countries, France, Cameroon and Vietnam, because of their
CC aberrant reactivities. The RNA was extracted, cDNA synthesised and PCR
CC amplified, cloned and genotyped. The 5'UR, Core/E1 and NS5B regions were
CC sequenced either directly or partially and used to classify the new
CC viruses into (sub)types based on comparison with known sequences.
CC The sequences were used to generate the peptides R96424-R96524. The
CC sequences can also be used to synthesise probes and primers for the
CC detection of HCV in a sample. The polypeptides can be used to detect
CC anti-HCV antibodies, for HCV typing or to prevent HCV infections.
SQ Sequence 340 BP; 85 A; 86 C; 94 G; 75 T;

Query Match 13.5%; Score 26.2; DB 1; Length 340;
Best Local Similarity 49.6%; Pred. No. 8.3;
Matches 67; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

```
Qy     36 AATGCCAGGTATGCAGACCGAGATCAAAACGCAGGCCACGAATCTTGCGGCGAATCTTTC 95
Db    304 ACTCTCAGCAATCACGACCAAGTCATCTCCGCAGACAAGGAATGATGGATTTTTAATTCC 245

Qy     96 CGCAGTCAGAGAAAGTGCCACAGCGACGCTGTCAGGGGAAATTAAAGGCCCGCAACTGGA 155
Db    244 GGCAGCTTTGGTAGCTGCCTTTGCTTTGATGTAACAGGTGATTGTGTTCCCGAAGCTGGT 185

Qy    156 AGATTTTCCCGCGCT 170
Db    184 GGGTAGCACCCCACT 170
```

Search completed: October 5, 1999, 17:59:10
Job time: 11392 sec

FT in the specification"
PN WO9517508-A2.
PD 29-JUN-1995.
PF 21-DEC-1994; CA0705.
PR 23-DEC-1993; US-172188.
PA (ALLX) ALLELIX BIOPHARMACEUTICALS INC.
PI Kamboj R, Nutt S;
DR WPI; 95-240670/31.
DR P-PSDB; R75882.
PT Identification of human CNS receptor ligand - and identification of
PT agents that modulate editing of human CNS receptors
PS Example 1; Fig 1; 59pp; English.
CC The genomic sequence (Q91230) coding for the glutamate (EAA)
CC receptor GluR2B (R75882) was obtd. following PCR amplification, and
CC was compared to the GluR2B cDNA sequence. A single nucleotide
CC difference in the coding region of the transmembrane II domain, at
CC position 2134, was identified. The genomic DNA included a G
CC (thereby encoding glutamine), and the cDNA an A (thereby encoding
CC arginine). The difference was a result of RNA editing. The edited
CC and unedited receptors are functionally distinct and appear to be
CC selectively expressed. CNS receptor genes subject to such editing
CC (see also Q91231 and Q91232) have value in screening cpds. for
CC potential therapeutic utility.
SQ Sequence 3407 BP; 998 A; 657 C; 844 G; 907 T;

Query Match 13.6%; Score 26.4; DB 1; Length 3407;
Best Local Similarity 47.6%; Pred. No. 15;
Matches 78; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

Qy 10 AGTGTAAGGACTCAGCCCCCCGTCATAATGCCAGGTATGCAGACCGAGATCAAAACGCAG 69
Db 2872 AGAATATTAACCCATCTTCCTCGCAGAATTCACAGAATTTTGCAACTTATAAGGAAGGTT 2931

Qy 70 GCCACGAATCTTGCGGCGAATCTTTCCGCAGTCAGAGAAAGTGCCACAGCGACGCTGTCA 129
Db 2932 ACAACGTATATGGCATCGAAAGTGTTAAAATTTAGGGGATGACCTTGAATGATGCCATGA 2991

Qy 130 GGGGAAATTAAAGGCCCGCAACTGGAAGATTTTCCCGCGCTGAT 173
Db 2992 GGAACAAGGCAAGGCTGTCAATTACAGGAAGTACTGGAGAAAAT 3035

RESULT 12
V54351/c
ID V54351 standard; DNA; 360 BP.
AC V54351;
DT 15-JAN-1999 (first entry)
DE Human coactivator gene.
KW Identification; gene expression; conserved motif; target; metastatic;
KW non-metastatic; non-cancerous tissue; nucleic acid analysis; ss.
OS Homo sapiens.
PN WO9839480-A1.
PD 11-SEP-1998.
PF 03-MAR-1998; U04094.
PR 03-MAR-1997; US-785522.
PA (HAQQ/) HAQQI T M.
PI Haqqi TM;
DR WPI; 98-506373/43.
PT Method of analysing nucleic acid in a sample - used for analysing
PT expressed genes, and distinguishing between metastatic,
PT non-metastatic and non-cancerous tissues
PS Example 3; Fig 7; 55pp; English.
CC A novel method has been developed of analysing a nucleic acid in a
CC sample. The method comprises: (a) providing: (i) a sample containing
CC nucleic acid; (ii) a first primer having a sequence of which at least a
CC portion is at least partially complementary to a natural common non-
CC coding sequence on a portion of the nucleic acid of the sample; (iii) a
CC second primer having a sequence of which at least a portion is at least
CC partially complementary to a restriction enzyme recognition sequence
CC present on a portion of the nucleic acid of the sample; and (iv) a
CC polymerase and polymerase chain reaction (PCR) reagents; (b) preparing
CC the nucleic acid from the sample under conditions so as to produce
CC amplifyable nucleic acid; (c) amplifying the nucleic acid with the first
CC and second primers, the polymerase and the PCR reagents under conditions
CC such that amplified product is generated; and (d) detecting the amplified
CC product. The method can be used for analysing expressed genes in
CC multiple samples, especially in human cancer cells. The method can also
CC be used to distinguish between metastatic and non-metastatic cancer
CC tissues as well as between normal and cancerous tissue. It can also be
CC used to detect drug resistance. The method can also be used to classify
CC and identify microorganisms. The present sequence represents a human
CC coactivator gene used in an example from the present invention.
SQ Sequence 360 BP; 50 A; 143 C; 96 G; 71 T;

Query Match 13.6%; Score 26.4; DB 1; Length 360;
Best Local Similarity 55.4%; Pred. No. 7.3;
Matches 51; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

Qy 72 CACGAATCTTGCGGCGAATCTTTCCGCAGTCAGAGAAAGTGCCACAGCGACGCTGTCAGG 131
Db 358 CACGAATGTTGCTGAGGTTGATCTGCCGCTCAGGAGGAGGCCCACCACGAGGCTGACCTC 299

Qy 132 GGAAATTAAAGGCCCGCAACTGGAAGATTTTC 163
Db 298 GGACAATGATGGCGCACCCTGAGAGGACCATC 267

RESULT 13
X14126/c
ID X14126 standard; DNA; 1237 BP.
AC X14126;
DT 31-MAR-1999 (first entry)
DE H. pylori GHPO 229 gene.
KW GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW peptic ulcer disease; ss.
OS Helicobacter pylori.
FH Key Location/Qualifiers
FT CDS 51. .1187
FT /*tag= a
PN WO9843478-A1.
PD 08-OCT-1998.
PF 01-APR-1998; U06371.
PR 29-JUL-1997; US-902615.
PR 01-APR-1997; US-833457.
PR 24-JUN-1997; US-881227.
PA (HUMA-) HUMAN GENOME SCI INC.
PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
PI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;
DR WPI; 98-542293/46.
DR P-PSDB; W98407.
PT New isolated Helicobacter polynucleotides - used to develop products
PT for the diagnosis, prevention and treatment of Helicobacter
PT infections and gastrointestinal diseases
PS Claim 1; Page 676-677; 2054pp; English.
CC This sequence represents a polynucleotide of the invention. It was
CC isolated from Helicobacter pylori and encodes a H.pylori GHPO protein.
CC The polypeptides can be used for preventing or treating Helicobacter
CC infections, and gastroduodenal diseases associated with these
CC infections, including acute, chronic, and atrophic gastritis, and peptic
CC ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used
CC for the production of antibodies. The products can also be used for
CC detection and diagnosis.
SQ Sequence 1237 BP; 313 A; 238 C; 287 G; 399 T;

Query Match 13.6%; Score 26.4; DB 1; Length 1237;
Best Local Similarity 50.8%; Pred. No. 11;
Matches 63; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

Qy 34 ATAATGCCAGGTATGCAGACCGAGATCAAAACGCAGGCCACGAATCTTGCGGCGAATCTT 93
Db 217 ATCATTGCATTAAAACTCTCCGTGATCCCAAGCCCTCCCATAAGTCCTGCTATGAGAGTC 158

Qy 94 TCCGCAGTCAGAGAAAGTGCCACAGCGACGCTGTCAGGGGAAATTAAAGGCCCGCAACTG 153

```
CC   encoding a polypeptide with the activity of interest or a related
CC   activity; and (v) isolating the hybrid DNA sequence identified in step
CC   (iv). This method provides for identification and isolation of sequences
CC   from microorganisms without having to cultivate and isolate the
CC   microorganism.
SQ   Sequence   212 BP;    42 A;    60 C;    60 G;    50 T;


  Query Match            13.7%;  Score 26.6;  DB 1;  Length 212;
  Best Local Similarity  66.7%;  Pred. No. 5.3;
  Matches   38;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy      83 CGGCGAATCTTTCCGCAGTCAGAGAAAGTGCCACAGCGACGCTGTCAGGGGAAATTA 139
           |||||   | |||||||  ||| ||   |    || ||||  |||| |||| |||||
Db     171 CGGCGGGGCATTCCGCACCCAGGGAGCATCTGCCAACGACTTTGTCTGGGGCAATTA 115


RESULT   9
Q54117
ID   Q54117 standard; cDNA; 3407 BP.
AC   Q54117;
DT   29-JUN-1994  (first entry)
DE   Human GluR2B receptor DNA.
KW   GluR2B receptor; glutamate receptor; AMPA-binding receptor;
KW   CNS; central nervous system; plasmid pBS/humGluR2B; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             315. .2966
FT                   /*tag=  a
FT   signal_peptide  315. .374
FT                   /*tag=  b
FT   mat_peptide     375. .2966
FT                   /*tag=  c
PN   EP-574257-A.
PD   15-DEC-1993.
PF   10-JUN-1993; 304500.
PR   10-JUN-1992; US-896437.
PR   10-JUN-1992; US-896611.
PR   10-JUN-1992; US-896612.
PA   (ALLX ) ALLELIX BIOPHARMACEUTICALS.
PA   (ELLI/) ELLIOTT C.
PA   (KAMB/).
PA   KAMBOJ R.
PA   (NUTT/) NUTT S L.
PI   Elliott C,  Kamboj R,  Nutt SL;
DR   WPI; 93-396903/50.
DR   P-PSDB; R45141.
PT   Polynucleotide(s) encoding AMPA-binding human glutamate receptors
PT   - used for providing prods. for use in screening assays or in
PT   detection and/or localisation
PS   Claim 1; Page 22-26; 101pp; English.
CC   The GluR2B receptor DNA is contained within plasmid pBS/humGluR2B
CC   (ATCC 75217) and may be expressed by a mammal recombinant cell.  The
CC   DNA can be used to obtain products for evaluating the human nervous
CC   system and particularly for assessing potentially therapeutic
CC   interactions between the AMPA-binding human excitatory amino acid
CC   receptors and natural and synthetic ligands.  It can also be used
CC   for detecting the presence and/or location of human GluR receptors
CC   e.g. in brain tissue.  The DNA was isolated from human brain cDNA
CC   libraries using rat GluR receptor gene probes.
SQ   Sequence   3407 BP;    997 A;    657 C;    845 G;    907 T;


  Query Match            13.6%;  Score 26.4;  DB 1;  Length 3407;
  Best Local Similarity  47.6%;  Pred. No. 15;
  Matches   78;  Conservative    0;  Mismatches   86;  Indels    0;  Gaps    0;

Qy      10 AGTGTAAGGACTCAGCCCCCCGTCATAATGCCAGGTATGCAGACCGAGATCAAAACGCAG 69
           ||  ||   || || |  ||   || ||| |   | ||   |      || |  | |
Db    2872 AGAATATTAACCCATCTTCCTCGCAGAATTCACAGAATTTTGCAACTTATAAGGAAGGTT 2931

Qy      70 GCCACGAATCTTGCGGCGAATCTTTCCGCAGTCAGAGAAAGTGCCACAGCGACGCTGTCA 129
           | ||| || | ||  ||||  | |    | | || | | |   |  || ||  | |
Db    2932 ACAACGTATATGGCATCGAAAGTGTTAAAATTTAGGGGATGACCTTGAATGATGCCATGA 2991

Qy     130 GGGGAAATTAAAGGCCCGCAACTGGAAGATTTTCCCGCGCTGAT 173
           ||   ||   ||||| |||| |  | ||  | |  | |   ||
Db    2992 GGAACAAGGCAAGGCTGTCAATTACAGGAAGTACTGGAGAAAAT 3035


RESULT  10
Q90296
ID   Q90296 standard; cDNA; 733 BP.
AC   Q90296;
DT   07-JAN-1996  (first entry)
DE   Human Bcl-2-associated protein BAG-1 cDNA.
KW   Cell death; apoptosis; bcl-2; BAG-1; BAP; athanogene-1; cancer;
KW   neurodegenerative disease; gene therapy; antisense; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             6. .716
FT                   /*tag=  a
PN   WO9513292-A1.
PD   18-MAY-1995.
PF   09-NOV-1994; U12904.
PR   12-NOV-1993; US-152485.
PA   (LJOL-) LA JOLLA CANCER RES FOUND.
PI   Reed JC,  Sato T,  Takayama S;
DR   WPI; 95-194030/25.
DR   P-PSDB; R74822.
PT   Nucleic acid encoding Bcl-2-associated proteins - used to develop
PT   prods. for modulating apoptosis in cells for the treatment of e.g.
PT   neurodegenerative disease or cancer.
PS   Claim 6; Page 59; 73pp; English.
CC   Nucleic acid encoding BAG-1 can be introduced into a cell to
CC   increase the ability of cell to survive, and antisense sequences can
CC   be introduced to decrease the ability of a cell to survive.  This
CC   cDNA sequence was cloned from a phage lambda Exlox expression vector
CC   library, which was screened using a recombinant Bcl-2 protein and an
CC   antibody to Bcl-2.  Mouse BAG-1 cDNA was used as a probe to
CC   screen a human breast cDNA library to obtain human BAG-1 cDNA.
SQ   Sequence   733 BP;    219 A;    148 C;    193 G;    173 T;


  Query Match            13.6%;  Score 26.4;  DB 1;  Length 733;
  Best Local Similarity  51.7%;  Pred. No. 9.2;
  Matches   60;  Conservative    0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy      19 ACTCAGCCCCCCGTCATAATGCCAGGTATGCAGACCGAGATCAAAACGCAGGCCACGAAT 78
           ||| | ||  ||   | | | | | |   | |||  || |   ||| ||  |  |  ||
Db     415 ACTGATCCTGCCAGAAAATTTCAAAGACAGTAGATTGAAAAGGAAAGGCTTGGTAAAAAA 474

Qy      79 CTTGCGGCGAATCTTTCCGCAGTCAGAGAAAGTGCCACAGCGACGCTGTCAGGGGA 134
           | | |  | || |  |  |||  || | ||||   |||     ||| |||| ||
Db     475 GGTTCAGGCATTCCTAGCCGAGTGTGACACAGTGGAGCAGAACATCTGCCAGGAGA 530


RESULT  11
Q91230
ID   Q91230 standard; DNA; 3407 BP.
AC   Q91230;
DT   10-JUL-1996  (first entry)
DE   Human GluR2B receptor (Q-587) gene.
KW   Glutamate receptor; EAA receptor; excitatory amino acid;
KW   CNS receptor; RNA editing; GluR2B; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             315. .2966
FT                   /*tag=  a
FT   signal_peptide  315. .374
FT                   /*tag=  b
FT   mat_peptide     375. .2963
FT                   /*tag=  c
FT   misc_difference 3068
FT                   /*tag=  d
FT                   /note= "base n at position 3068 is not identified
```

```
RESULT   6
Q50168
ID   Q50168 standard; DNA; 8298 BP.
AC   Q50168;
DT   04-MAY-1994  (first entry)
DE   p190 protein gene.
KW   p190; phosphoprotein; GTPase activating protein; GAP; n-chimerin;
KW   mitogenically-stimulated cells; tyrosine kinase-transformed cells;
KW   GAP-associated protein; homology; superfamily; signal transduction;
KW   transcription repressor; GRF-1; BCR; breakpoint cluster region gene;
KW   p21ras; effector; GTPase; mitogenic; ds.
OS   Rattus rattus.
FH   Key             Location/Qualifiers
FT   5'utr           1. .730
FT                   /*tag= a
FT   cds             731. .5272
FT                   /*tag= b
FT                   /product= p190 protein
FT   3'utr           5273. .8298
FT                   /*tag= c
PN   WO9320201-A.
PD   14-OCT-1993.
PF   31-MAR-1993; U03076.
PR   31-MAR-1992; US-861207.
PA   (WHED ) WHITEHEAD INST BIOMEDICAL RES.
PI   Settleman JE,  Weinberg RA;
DR   WPI; 93-336909/42.
DR   P-PSDB; R43253.
PT   New GAP-associated protein P190 - which can be inhibited to
PT   interfere with RAS oncogene(s) in pathogenesis of malignancies
PS   Claim 4; Page 42-55; 95pp; English.
CC   This sequence encodes the p190 protein.  p190 is a phosphoprotein
CC   which is tightly bound to GTPase activating protein (GAP) in
CC   mitogenically-stimulated and tyrosine kinase-transformed cells.
CC   p190 is GAP-associated protein.  p190 has three distinct domains,
CC   each of which exhibits homology to a previously described sequence.
CC   Towards the amino terminal end of p190, a domain spanning 201 amino
CC   acids exhibits significant sequence similarity to all members of the
CC   GTPase superfamily.  The most striking similarity in the predicted
CC   amino acid sequence of p190, is to a 95 kD protein encoded by a human
CC   cDNA which is reported to function as a transcription repressor.  The
CC   reported amino acid sequence of the represssor protein, GRF-1, is 95%
CC   identical over a 778 amino acid fragment of p190.  Towards the carboxy
CC   terminal of p190 there is a region of 161 residues which shows homology
CC   with two proteins which are involved in signal transduction.  It has
CC   been suggested that p21ras acts as a regulatory subunit of p190
CC   protein, which acts as the p21ras effector and which releases mitogenic
CC   signals when prompted to do so by activated GTP-bound p21ras.  p190 may
CC   also, acting via GAP, transduce signals from p21ras to the nucleus,
CC   affecting expression of specific cellular genes.
SQ   Sequence   8298 BP;    2180 A;    2086 C;    2039 G;    1993 T;


  Query Match             14.0%;  Score 27.2;  DB 1;  Length 8298;
  Best Local Similarity   58.8%;  Pred. No. 11;
  Matches   47;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy      3 GGTTACAAGTGTAAGGACTCAGCCCCCCGTCATAATGCCAGGTATGCAGACCGAGATCAA 62
          ||  ||||||| |  ||||| || |  | | |||  || |||   |||||  |  | |||
Db   2851 GGACACAAGTGGAGAGACTCTGCACAGCTTAATACAGCAAGGCCAGCAGATTGCTAGCAA 2910

Qy     63 AACGCAGGCCACGAATCTTG 82
          |   |||        |||||
Db   2911 ACTTCAGTGTGTCTTTCTTG 2930


RESULT   7
N60971
ID   N60971 standard; DNA; 2615 BP.
AC   N60971;
DT   28-OCT-1991  (first entry)
DE   Sequence encoding homoserine dehydrogenase and homoserine kinase.
KW   L-threonine; L-isoleucine; Ile; Thr; ss.
OS   Corynebacterium glutamicum.
FH   Key             Location/Qualifiers
FT   cds             322. .1560
FT                   /*tag= a
FT                   /label= Homoserine dehydrogenase
FT   cds             1571. .2497
FT                   /*tag= b
FT                   /label= Homoserine kinase
PN   EP-204326-A.
PD   10-DEC-1986.
PF   04-JUN-1986; 107595.
PR   05-JUN-1985; JP-121674.
PR   31-JAN-1986; JP-019801.
PR   02-APR-1986; JP-076298.
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
PI   Katsumata R, Mizukami T, Yokoi M, Kikuchi Y, Oka T;
DR   WPI; 86-327151/50.
DR   P-PSDB; P61449 and P61775.
PT   Prodn. of L-threonine and L-isoleucine - by cultivating
PT   corynebacterium or brevibacterium strain transformed with
PT   recombinant DNA.
PS   Claim 5; Table 3; 34pp; English.
CC   The sequence expresses two enzymes useful in high yield production of
CC   L-isoleucine and L-threonine for medicine and food, for human and
CC   animal consumption. The products may be expressed from a transformed
CC   host micro-organism.
SQ   Sequence   2615 BP;    570 A;    667 C;    726 G;    652 T;


  Query Match             13.7%;  Score 26.6;  DB 1;  Length 2615;
  Best Local Similarity   52.2%;  Pred. No. 12;
  Matches   59;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy     81 TGCGGCGAATCTTTCCGCAGTCAGAGAAAGTGCCACAGCGACGCTGTCAGGGGAAATTAA 140
          ||| ||| | |||||||  | | |   ||  ||||  | |  |    | | |  ||||
Db   2347 TGCAGCGTACCTTTCCGGTGCCGGCCCAACCGCCATGGTGCTGTCCACTGAGCCAATTCC 2406

Qy    141 AGGCCCGCAACTGGAAGATTTTCCCGCGCTGATCAAACAGGCGAGTCTGGATG 193
          || |  |    ||||||||  ||  | |     ||   ||| |  |  |  ||
Db   2407 AGACAAGGTTTTGGAAGATGCTCGTGAGTCTGGCATTAAGGTGCTTGAGCTTG 2459


RESULT   8
V15054/c
ID   V15054 standard; DNA; 212 BP.
AC   V15054;
DT   19-MAY-1998  (first entry)
DE   Xylanase activity DNA sequence from a cow rumen sample SEQ ID NO:44.
KW   Bacillus sp.; xylanase; glycosyl hydrolase family 11; isolation;
KW   microorganism; identification; hybrid DNA; ss.
OS   Synthetic.
OS   Unidentified.
PN   WO9743409-A2.
PD   20-NOV-1997.
PF   12-MAY-1997; DK0216.
PR   10-MAY-1996; DK-000562.
PA   (NOVO ) NOVO-NORDISK AS.
PI   Dalboge H, Diderichsen B, Kauppinen S, Sandal T;
DR   WPI; 98-008878/01.
PT   Isolating novel DNA sequences from microorganisms - without the need
PT   for culturing the microorganism
PS   Example 2; Page 44; 72pp; English.
CC   The present sequence represents a novel sequence with xylanase activity
CC   from an example of the present invention. The present invention
CC   describes a novel method for providing a novel DNA sequence encoding a
CC   polypeptide from a microorganism with an activity of interest. The
CC   method comprises: (i) PCR amplification of the DNA with PCR primers with
CC   homology to (a) known gene(s) encoding a polypeptide with an activity of
CC   interest; (ii) linking the obtained PCR product of a 5' structural gene
CC   sequence and a 3' structural gene sequence; (iii) expressing the
CC   resulting hybrid DNA sequence; (iv) screening for hybrid DNA sequences
```

```
CC   less promoters.  In fact, this region of the AAVS1 sequence lacks a
CC   TATA-like element.  Other landmarks within the CpG island are SP1-
CC   like binding sites, a CRE and nine UBF-1 sequences.  Although the
CC   identity of the protein encoded by this gene is not certain, the
CC   presence of these elements indicate it is a regulated protein, and
CC   that the gene is thus not just a constitutive housekeeping gene.
SQ   Sequence   4067 BP;    552 A;    1383 C;    1168 G;    962 T;


  Query Match             15.7%;  Score 30.4;  DB 1;  Length 4067;
  Best Local Similarity   51.5%;  Pred. No. 0.76;
  Matches   70;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy      55 GAGATCAAAACGCAGGCCACGAATCTTGCGGCGAATCTTTCCGCAGTCAGAGAAAGTGCC 114
           || |   |  | ||||    ||| || || | |||||      ||  |||||| |||| |
Db    3465 GAAACACAGCCCCAGGTGGAGAAACTGGCCGGGAATCAAGAGTCACCCAGAGACAGTGAC 3407

Qy     115 ACAGCGACGCTGTCAGGGGAAATTAAAGGCCCGCAACTGGAAGATTTTCCCGCGCTGATC 174
           | |  | ||||      |      |||    || |   || | |   | |  |||  |
Db    3406 CAACCATCCCTGTTTTCCTAGGACTGAGGGTTTCAGCGCTAAAACTAGGCTGTCCTGGGC 3347

Qy     175 AAACAGGCGAGTCTGG 190
           ||||||   |  ||||
Db    3346 AAACAGCATAAGCTGG 3331


RESULT   4
T98709/c
ID   T98709 standard; DNA; 874 BP.
AC   T98709;
DT   09-NOV-1998  (first entry)
DE   DNA encoding a S. pneumoniae amino acid permease ROCE.
KW   Streptococcus pneumoniae protein; genetic immunisation; antagonist;
KW   immunological response; inoculation; antibody production; inhibitor;
KW   T cell immune response; antimicrobial compound; bacterial adhesion;
KW   extracellular matrix protein; protein-mediated cell invasion; wound;
KW   pathogenesis; ss.
OS   Streptococcus pneumoniae.
FH   Key             Location/Qualifiers
FT   CDS             537. .866
FT                   /*tag=  a
PN   WO9743303-A1.
PD   20-NOV-1997.
PF   14-MAY-1997; U07950.
PR   14-MAY-1996; US-017670.
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PA   (SMIK ) SMITHKLINE BEECHAM PLC.
PI   Black MT, Hodgson JE, Knowles DJC, Nicholas RO,
PI   Stodola RK;
DR   WPI; 98-008793/01.
DR   P-PSDB; W38665.
PT   Novel Streptococcus pneumoniae proteins and related DNA - useful for
PT   diagnosing anti-microbial agents for treatment of bacterial
PT   infections
PS   Claim 4; Page 212; 483pp; English.
CC   This sequence encodes a Streptococcus pneumoniae protein that (based on
CC   homology with a bacillus subtilis protein) is a amino acid permease
CC   ROCE, and represents a DNA sequence of the invention.
CC   The DNA sequences were isolated from Streptococcus pneumoniae strain
CC   0100993 (NCIMB 40794). The Streptococcus pneumoniae proteins of the
CC   invention can be used to identify compounds which interact with and
CC   inhibit or activate the activity of the proteins. Antagonists can be
CC   used to treat diseases caused by S. pneumoniae proteins, through genetic
CC   immunisation. They can also be used to induce an immunological response
CC   in a mammal by inoculation with the S. pneumoniae proteins or delivery
CC   of the encoding nucleic acids in a vector adequate to produce antibody
CC   and/or T cell immune responses to protect the animal from disease. The
CC   proteins can also be used to identify antimicrobial compounds which are
CC   capable of inhibiting their bioactivity. In particular the proteins of
CC   the invention can be used to prevent adhesion of bacteria to mammalian
CC   extracellular matrix proteins on in-dwelling devices or in wounds, to
CC   block protein-mediated mammalian cell invasion, and to block the normal
CC   progression of pathogenesis in infections initiated other than by the
CC   implantation of in-dwelling devices or other surgical techniques.
SQ   Sequence   874 BP;    199 A;    168 C;    195 G;    312 T;


  Query Match             14.9%;  Score 29;  DB 1;  Length 874;
  Best Local Similarity   50.4%;  Pred. No. 1.3;
  Matches   71;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy       5 TTACAAGTGTAAGGACTCAGCCCCCCGTCATAATGCCAGGTATGCAGACCGAGATCAAAA 64
           | | | | |  ||| |  ||  |  ||  |||    | |  ||  ||  | ||||    |
Db     643 TCATACGCGATAGGGCATAGGTCATCGAAATACAAACTGTAATCAAGGTCAAGATAGCCA 584

Qy      65 CGCAGGCCACGAATCTTGCGGCGAATCTTTCCGCAGTCAGAGAAAGTGCCACAGCGACGC 124
           |  | | |||  |  | || ||  | ||   | ||      |  ||| || |  || |||
Db     583 CTAATGACACATAGTTGGCTGCCCAACTAATCCCAACACTACGAAGGGCAAAGGCAACGG 524

Qy     125 TGTCAGGGGAAATTAAAGGCC 145
           |||  |  | |||||  | |
Db     523 CATCATCGACATTTAAATGAC 503


RESULT   5
Q11850
ID   Q11850 standard; DNA; 3505 BP.
AC   Q11850;
DT   31-JUL-1991  (first entry)
DE   Glutamate receptor 2.
KW   Glutamate receptor 2; probe; ligand; drug screening; ss.
OS   Rattus rattus.
FH   Key             Location/Qualifiers
FT   cds             316. .2964
FT                   /*tag=  a
FT                   /product= GR2
FT   signal_peptide  316. .378
FT                   /*tag=  b
FT   mat_peptide     379. .2964
FT                   /*tag=  c
PN   WO9106648-A.
PD   16-MAY-1991.
PF   25-OCT-1990; U06153.
PR   27-OCT-1989; US-428116.
PA   (SALK ) SALK INST FOR BIOL STUD.
PI   Heinemann SF, Boulter JR, Hollmann M, Bettler B, Jensen JE;
DR   WPI; 91-164197/22.
DR   P-PSDB; R11990.
PT   Glutamate receptors - used to screen for functional ligands and
PT   identify and isolate further receptors
PS   Disclosure; Fig 3; 109pp; English.
CC   GluR2 cDNA was isolated from a rat forebrain cDNA using a
CC   low-stringency screening protocol and a radiolabelled fragment of the
CC   GluR1 cDNA as probe.  The cDNA is deposited (ATCC 68132).
CC   The gene and protein can be used in drug screening, to
CC   determine whether a substance is a functional ligand for the
CC   receptor by monitoring ion channel activity.
CC   See also Q11849-855.
SQ   Sequence   3505 BP;    1032 A;    696 C;    869 G;    908 T;


  Query Match             14.4%;  Score 28;  DB 1;  Length 3505;
  Best Local Similarity   48.2%;  Pred. No. 4.5;
  Matches   79;  Conservative    0;  Mismatches   85;  Indels    0;  Gaps    0;

Qy      10 AGTGTAAGGACTCAGCCCCCCGTCATAATGCCAGGTATGCAGACCGAGATCAAAACGCAG 69
           ||  ||   || || |  ||   || ||| ||  | ||  |       || | | |
Db    2873 AGAATATTAACCCATCTTCCTCGCAGAATTCCCAGAATTTTGCAACTTATAAGGAAGGTT 2932

Qy      70 GCCACGAATCTTGCGGCGAATCTTTCCGCAGTCAGAGAAAGTGCCACAGCGACGCTGTCA 129
           | ||| || | ||  |||   | |    | | || | | |  |   || || |   | |
Db    2933 ACAACGTATATGGCATCGAGAGTGTTAAAATTTAGGGGATGACCTTGAGTGATGTCATGA 2992

Qy     130 GGGGAAATTAAAGGCCCGCAACTGGAAGATTTTCCCGCGCTGAT 173
           || | ||   |||||   ||| |  | ||  | |  | |   ||
Db    2993 GGAGCAAGGCAAGGCTGTCAATTACAGGAAGTACTGGAGAAAAT 3036
```

```
RESULT   2
T67035
ID   T67035 standard; DNA; 3441 BP.
AC   T67035;
DT   04-AUG-1997  (first entry)
DE   Salmonella virulence factor SspBCDA DNA region.
KW   Salmonella secreted protein; SspB; SspC; SspD; SspA;
KW   bacterial-mediated endocytosis; diagnosis; therapy; vaccine;
KW   attenuation; virulence; ss.
OS   Salmonella typhimurium.
FH   Key             Location/Qualifiers
FT   cds             1. .870
FT                   /*tag= a
FT                   /product= SspB (incomplete)
FT   cds             898. .2127
FT                   /*tag= b
FT                   /product= SspC
FT   cds             2198. .3228
FT                   /*tag= c
FT                   /product= SspD
FT   cds             3247. .3441
FT                   /*tag= d
FT                   /product= SspA (incomplete)
PN   WO9718225-A1.
PD   22-MAY-1997.
PF   14-NOV-1996; U18504.
PR   14-NOV-1995; US-006733.
PA   (GEHO ) GEN HOSPITAL CORP.
PI   Miller SI;
DR   WPI; 97-289217/26.
DR   P-PSDB; W15268-71.
PT   New isolated Salmonella secreted proteins and related genes - used
PT   to develop products for the detection, treatment or prevention of
PT   Salmonella infections
PS   Claim 11; Fig 28; 95pp; English.
CC   A DNA molecule (T67035) comprises the SspBCDA locus (truncated at
CC   the 5' and 3' ends) including the partial sspB, sspC, sspD and
CC   partial sspA sequences (see also T67028-31) coding for Salmonella
CC   typhimurium secreted proteins SspB, SspC, SspD and SspA (W15268-
CC   70).  To identify the ssp genes, S. typhimurium mutants were
CC   investigated in invasion assays on HEp-2 epithelial cells.  2
CC   Ssp mutants with transposon insertions located between spaT and
CC   prgH (see also T67036) were identified.  One mutant lacked Ssp of
CC   87, 42 and 36 kDa.  The invasiveness of this mutant implicated the
CC   42 and 36 kDa Ssp in invasion.  DNA sequencing of this region
CC   identified the 4 ssp (partial) genes.  The ssp genes can be used in
CC   methods for diagnosing Salmonella infection and in methods of
CC   targetting an antigen to epithelial cells and of inducing a
CC   cytotoxic T cell immune response in a mammal.  Ssp gene mutation
CC   can be used to create attenuated Salmonella strains useful as live
CC   vaccines.
SQ   Sequence   3441 BP;    978 A;    761 C;    925 G;    777 T;


  Query Match             94.3%;  Score 183;  DB 1;  Length 3441;
  Best Local Similarity   99.5%;  Pred. No. 1.5e-51;
  Matches  194;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy       1 ATGGTTACAAGTGTAAGGACTCAG-CCCCCCGTCATAATGCCAGGTATGCAGACCGAGAT 59
           |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db    3247 ATGGTTACAAGTGTAAGGACTCAGCCCCCCCCGTCATAATGCCAGGTATGCAGACCGAGAT 3306

Qy      60 CAAAACGCAGGCCACGAATCTTGCGGCGAATCTTTCCGCAGTCAGAGAAAGTGCCACAGC 119
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3307 CAAAACGCAGGCCACGAATCTTGCGGCGAATCTTTCCGCAGTCAGAGAAAGTGCCACAGC 3366

Qy     120 GACGCTGTCAGGGGAAATTAAAGGCCCGCAACTGGAAGATTTTCCCGCGCTGATCAAACA 179
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3367 GACGCTGTCAGGGGAAATTAAAGGCCCGCAACTGGAAGATTTTCCCGCGCTGATCAAACA 3426

Qy     180 GGCGAGTCTGGATGC 194
           |||||||||||||||
Db    3427 GGCGAGTCTGGATGC 3441


RESULT   3
Q63192/c
ID   Q63192 standard; DNA; 4067 BP.
AC   Q63192;
DT   18-NOV-1994  (first entry)
DE   Human adeno-associated virus integration locus, AAVS1.
KW   Adeno-associated virus; AAV; integration locus; CpG island;
KW   SP1-like binding site; cAMP response element; CRE;
KW   upstream binding factor 1; UBF-1; minisatellite;
KW   probe; gene therapy; promoter; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             1616. .2318
FT                   /*tag= a
FT                   /note= "position indicated in Fig differs from
FT                   the position described in the text;
FT                   see CC"
FT   protein_bind    1. .900
FT                   /*tag= b
FT                   /label= GC_rich_region
FT                   /note= "GC-rich region indicative of a CpG island;
FT                   within the CpG island are SP1-like binding
FT                   sites (positions 293-300, 355-375 and
FT                   422-433, according to text), a cAMP response
FT                   element (see tag f), and nine upstream binding
FT                   factor 1 (UBF-1) sequences (CGGCC or
FT                   GGCCG) (positions 87, 347, 455, 497, 585,
FT                   587, 603, 689 and 690 according to the text);
FT                   see CC"
FT   protein_bind    2130
FT                   /*tag= c
FT                   /note= "additional UBF-1; position according to
FT                   text; see CC"
FT   protein_bind    3782
FT                   /*tag= d
FT                   /note= "additional UBF-1; position accoriding to
FT                   text; see CC"
FT   protein_bind    3441
FT                   /*tag= e
FT                   /note= "additional UBF-1; position accoriding to
FT                   text; see CC"
FT   misc_signal     278. .285
FT                   /*tag= f
FT                   /label= CRE
FT                   /note= "cAMP response element; see CC"
FT   misc_feature    3655. .4021
FT                   /*tag= g
FT                   /label= minisatellite_sequence
FT                   /note= "chromosome 19 specific;
FT                   position indicated in Fig differs
FT                   from the position described in the text;
FT                   see CC"
PN   EP-592836-A.
PD   20-APR-1994.
PF   16-SEP-1993; 114941.
PR   17-SEP-1992; US-947127.
PA   (AMCY ) AMERICAN CYANAMID CO.
PI   Berns KI,  Kotin RM,  Linden RM;
DR   WPI; 94-127741/16.
PT   New nucleic acid corresponding to human adeno-associated virus
PT   integration site - useful e.g., as probe to confirm targetted
PT   integration of adeno-associated virus vectors in gene therapy
PS   Claim 3; Page 7-9; 20pp; English.
CC   NB: Positions of features indicated in the Fig (Fig 1) differ
CC   from positions described in the text.  No features are indicated
CC   in the sequence of the Sequence Listing.
CC   The novel promoter/gene is characterised by a region of high
CC   G + C conc. in nucleotides 1-900, indicating the presence
CC   of a CpG island.  CpG have been previously identified in the
CC   literature as being associated with the 5' ends of genes and TATA-
```

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 1999, 17:59:04 ; Search time 223.92 Seconds
(without alignments)
216.762 Million cell updates/sec

Title: US-09-068-804-4
Perfect score: 194
Sequence: 1 ATGGTTACAAGTGTAAGGAC.....AAACAGGCGAGTCTGGATGC 194

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		194	100.0	194	1	T67031	Salmonella secrete
2		183	94.3	3441	1	T67035	Salmonella virulen
c 3		30.4	15.7	4067	1	Q63192	Human adeno-associ
c 4		29	14.9	874	1	T98709	DNA encoding a S.
5		28	14.4	3505	1	Q11850	Glutamate receptor
6		27.2	14.0	8298	1	Q50168	p190 protein gene.
7		26.6	13.7	2615	1	N60971	Sequence encoding
c 8		26.6	13.7	212	1	V15054	Xylanase activity
9		26.4	13.6	3407	1	Q54117	Human GluR2B recep
10		26.4	13.6	733	1	Q90296	Human Bcl-2-associ
11		26.4	13.6	3407	1	Q91230	Human GluR2B recep
c 12		26.4	13.6	360	1	V54351	Human coactivator
c 13		26.4	13.6	1237	1	X14126	H. pylori GHPO 229
c 14		26.2	13.5	280	1	Q78090	Hepatitis C virus
c 15		26.2	13.5	340	1	T27985	Hepatitis C virus
c 16		26	13.4	1196	1	Q37512	E. coli repeat 2 (
c 17		26	13.4	1365	1	X33776	S. aureus coding s
18		25.8	13.3	15239	1	T33536	BCG deletion regio
19		25.6	13.2	3685	1	N81712	Clone contg. hom-t
c 20		25.6	13.2	2160	1	Q66641	Mouse Activin rece
c 21		25.6	13.2	9034	1	T04247	Hepatitis GB virus
c 22		25.6	13.2	9122	1	T42920	GB-C viral genomic
c 23		25.6	13.2	9126	1	T45828	Hepatitis GB virus
c 24		25.6	13.2	9126	1	T96447	Fragment of HGBV N
c 25		25.6	13.2	9126	1	V39074	Hepatitis GB virus
c 26		25.6	13.2	1008	1	V39076	Hepatitis GB virus
27		25.6	13.2	1687	1	V84559	Human secreted pro
c 28		25.4	13.1	12283	1	Q10353	Nucleotide sequenc
c 29		25.4	13.1	1174	1	Q37510	E. coli repeat 1.
30		25.4	13.1	4040	1	T59731	Human interleukin-
c 31		25.2	13.0	110000	1	V30458_0	Rhizobium species
c 32		25.2	13.0	110000	1	V30459_0	Rhizobium species
33		25	12.9	4854	1	Q28399	Human nestin gene.
34		25	12.9	4852	1	Q70448	Human nestin gene
35		25	12.9	2363	1	T58313	Oyster pearl prote
c 36		25	12.9	5336	1	T90522	Arabidopsis male f
c 37		25	12.9	211	1	V15061	Xylanase activity
c 38		25	12.9	8752	1	V31189	E. coli J96 pathog
39		25	12.9	26016	1	V57272	Human flavin-conta
c 40		25	12.9	2919	1	V80687	Hampshire porcine
c 41		25	12.9	738	1	X39812	Gastric cancer ass
42		24.8	12.8	219	1	N92320	Murine autonomousl
43		24.8	12.8	1834	1	Q44495	Noggin clone A3. N
44		24.8	12.8	3900	1	Q91660	Drosophila patched
c 45		24.8	12.8	5467	1	T43685	Medium chain-speci

ALIGNMENTS

RESULT 1
T67031
ID T67031 standard; DNA; 194 BP.
AC T67031;
DT 04-AUG-1997 (first entry)
DE Salmonella secreted protein SspA DNA.
KW Salmonella secreted protein; SspA; bacterial-mediated endocytosis;
KW diagnosis; therapy; vaccine; attenuation; virulence; ss.
OS Salmonella typhimurium.
FH Key Location/Qualifiers
FT cds 1. .194
FT /*tag= a
FT /product= SspA (incomplete)
FT /note= "the translated sequence does not fully
FT correspond to the SspA amino acid sequence
FT given in Fig 23"
PN WO9718225-A1.
PD 22-MAY-1997.
PF 14-NOV-1996; U18504.
PR 14-NOV-1995; US-006733.
PA (GEHO) GEN HOSPITAL CORP.
PI Miller SI;
DR WPI; 97-289217/26.
DR P-PSDB; W15271.
PT New isolated Salmonella secreted proteins and related genes - used
PT to develop products for the detection, treatment or prevention of
PT Salmonella infections
PS Claim 9; Fig 22; 95pp; English.
CC A DNA molecule (T67031) codes for (incomplete) Salmonella secreted
CC protein SspA (W15271), a virulence factor able to translocate into
CC human epithelial cells. To identify Ssp genes, S. typhimurium
CC mutants were examined in invasion assays on HEp-2 epithelial cells.
CC 2 Ssp mutants with transposon insertions located between spaT and
CC prgH were identified. Of these, one mutant lacked Ssp of 87,
CC 42 and 36 kDa. The invasiveness of this mutant implicated the
CC 42 and 36 kDa Ssp in invasion. DNA sequencing of this region (see
CC also T67035) identified 2 complete and 2 partial open reading
CC frames, designated sspB, sspC (T67029), sspD (T67030) and sspA
CC (T67031). These nucleic acids can be used in methods for diagnosing
CC Salmonella infection and in methods of targetting an antigen to
CC epithelial cells and of inducing a cytotoxic T cell immune response
CC in a mammal. Ssp gene mutation can be used to create attenuated
CC Salmonella strains useful as live vaccines.
SQ Sequence 194 BP; 55 A; 51 C; 53 G; 35 T;

Query Match 100.0%; Score 194; DB 1; Length 194;
Best Local Similarity 100.0%; Pred. No. 1.3e-55;
Matches 194; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 ATGGTTACAAGTGTAAGGACTCAGCCCCCCGTCATAATGCCAGGTATGCAGACCGAGATC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATGGTTACAAGTGTAAGGACTCAGCCCCCCGTCATAATGCCAGGTATGCAGACCGAGATC 60

Qy      61 AAAACGCAGGCCACGAATCTTGCGGCGAATCTTTCCGCAGTCAGAGAAAGTGCCACAGCG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 AAAACGCAGGCCACGAATCTTGCGGCGAATCTTTCCGCAGTCAGAGAAAGTGCCACAGCG 120

Qy     121 ACGCTGTCAGGGGAAATTAAAGGCCCGCAACTGGAAGATTTTCCCGCGCTGATCAAACAG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 ACGCTGTCAGGGGAAATTAAAGGCCCGCAACTGGAAGATTTTCCCGCGCTGATCAAACAG 180

Qy     181 GCGAGTCTGGATGC 194
          ||||||||||||||
Db     181 GCGAGTCTGGATGC 194
```

```
  Matches   51;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy    102 CAGAGAAAGTGCCACAGCGACGCTGTCAGGGGAAATTAAAGGCCCGCAACTGGAAGATTT 161
          ||||     | |  | | || | ||     | |  || | || || || ||||  |  ||
Db   5207 CAGATGGTCTACACCTGAGATGGTGGATCCGCAGCTTGACGGTCCACAGCTGGCCGCATT 5266

Qy    162 TCCCGCGCTGATCAAACAGGCGAGTCTGGATGC 194
            | ||  || || ||| || , ||  | |||||
Db   5267 GGCTGCCGTGGTCGAACTGGGCAGCTTCGATGC 5299


RESULT  14
US-08-580-038-26/c
; Sequence 26, Application US/08580038
; Patent No. 5807670
;  GENERAL INFORMATION:
;    APPLICANT:  Muerhoff, A. S.
;    APPLICANT:  Simons, J. N.
;    APPLICANT:  Birkenmeyer, L.
;    APPLICANT:  Leary, T. P.
;    APPLICANT:  Erker, J. C.
;    APPLICANT:  Desai, S. M
;    APPLICANT:  Mushahwar, I. K.
;    APPLICANT:  Chalmers, M.
;    APPLICANT:  Dawson, G. J.
;    TITLE OF INVENTION:  Detection of Hepatitis GB Virus
;    TITLE OF INVENTION:  Genotypes
;    NUMBER OF SEQUENCES:  89
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  100 Abbott Park Rd
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  USA
;      ZIP:  60064
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/580,038
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Porembski, Priscilla E.
;      REGISTRATION NUMBER:  33,207
;      REFERENCE/DOCKET NUMBER:  5793.US.01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  708-937-0378
;      TELEFAX:  708-938-2623
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9126 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-580-038-26


  Query Match             13.2%;  Score 25.6;  DB 3;  Length 9126;
  Best Local Similarity   50.8%;  Pred. No. 15;
  Matches   61;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy     12 TGTAAGGACTCAGCCCCCCGTCATAATGCCAGGTATGCAGACCGAGATCAAAACGCAGGC 71
          |||| | || || |   |||||   | |  ||  ||||| |||  |  || | || |  |
Db   2104 TGTATGAACCCATCAGGCCGTCCCGAAGAGAGCCATGCAAACCCCGGGCAGACCGGATTC 2045

Qy     72 CACGAATCTTGCGGCGAATCTTTCCGCAGTCAGAGAAAGTGCCACAGCGACGCTGTCAGG 131
          | |  | |   | ||        |||| |    |||   |   | | |  | | || |||
Db   2044 CGCCGACCCATCAGCTCTGAACACCGCCGCACCAGAGGTTCGTAAAACCCCCCAGTGAGG 1985


RESULT  15
US-08-639-857-3/c
; Sequence 3, Application US/08639857
; Patent No. 5955318
;  GENERAL INFORMATION:
;    APPLICANT:  Simons, J. N.
;    APPLICANT:  Desai, S. M
;    APPLICANT:  Mushahwar, I. K.
;    TITLE OF INVENTION:  REAGENTS AND METHODS USEFUL FOR CONTROLLING THE
;    TITLE OF INVENTION:  TRANSLATION OF HEPATITIS GB PROTEINS
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  100 Abbott Park Rd
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  USA
;      ZIP:  60064
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/639,857
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Porembski, Priscilla E.
;      REGISTRATION NUMBER:  33,207
;      REFERENCE/DOCKET NUMBER:  5793.US.P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  708-937-0378
;      TELEFAX:  708-938-2623
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9126 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-639-857-3


  Query Match             13.2%;  Score 25.6;  DB 4;  Length 9126;
  Best Local Similarity   50.8%;  Pred. No. 15;
  Matches   61;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy     12 TGTAAGGACTCAGCCCCCCGTCATAATGCCAGGTATGCAGACCGAGATCAAAACGCAGGC 71
          |||| | || || |   |||||   | |  ||  ||||| |||  |  || | || |  |
Db   2104 TGTATGAACCCATCAGGCCGTCCCGAAGAGAGCCATGCAAACCCCGGGCAGACCGGATTC 2045

Qy     72 CACGAATCTTGCGGCGAATCTTTCCGCAGTCAGAGAAAGTGCCACAGCGACGCTGTCAGG 131
          | |  | |   | ||        |||| |    |||   |   | | |  | | || |||
Db   2044 CGCCGACCCATCAGCTCTGAACACCGCCGCACCAGAGGTTCGTAAAACCCCCCAGTGAGG 1985


Search completed: October  5, 1999, 18:01:06
Job time: 11506 sec
```

```
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/375,241
;       FILING DATE:
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/738,800
;       FILING DATE:  31-JUL-1991
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Granahan, Patricia
;       REGISTRATION NUMBER:  32,227
;       REFERENCE/DOCKET NUMBER:  GTR90-04
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617-861-6240
;       TELEFAX:  617-861-9540
;   INFORMATION FOR SEQ ID NO:  7:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1196 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
US-08-375-241-7


  Query Match             13.4%;  Score 26;  DB 1;  Length 1196;
  Best Local Similarity   46.9%;  Pred. No. 5;
  Matches   69;  Conservative    3;  Mismatches   75;  Indels    0;  Gaps    0;

Qy       1 ATGGTTACAAGTGTAAGGACTCAGCCCCCCGTCATAATGCCAGGTATGCAGACCGAGATC 60
           | |||  ||  | |  | |||:   :  |      ||  |   |  | ||||| ||| |
Db     634 AAGGTGGCACATATCCGCACTYGCNHAACATATGGAACCCGGCGGCTCCAGACGGAGCTG 575

Qy      61 AAAACGCAGGCCACGAATCTTGCGGCGAATCTTTCCGCAGTCAGAGAAAGTGCCACAGCG 120
             |  | | | |  |   |||  | ||     |     |||   | |||   |  ||:
Db     574 GCAGAGAATGGCATCATCGTTGGCGTGACCGACTGGCACGTCTTCGTAAGGAGCTAAGST 515

Qy     121 ACGCTGTCAGGGGAAATTAAAGGCCCG 147
           ||||||| ||  ||||   |||  | |
Db     514 ACGCTGTAAGCAGAAACGCAAGTTCAG 488


RESULT  12
PCT-US92-06617A-7/c
; Sequence 7, Application PC/TUS9206617A
;  GENERAL INFORMATION:
;    APPLICANT:  Parodos, Kyriaki
;    APPLICANT:  McCarty, Janice
;    TITLE OF INVENTION:  Nucleic Acid Probes for the Detection of
;    TITLE OF INVENTION:  Shigella
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Amoco Corporation
;      STREET:  200 East Randolph Drive, P.O. Box 87703
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  U.S.A.
;      ZIP:  60680
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/06617A
;      FILING DATE:  19920728
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/738,800
;      FILING DATE:  31-JUL-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Galloway, Norval B.
;      REGISTRATION NUMBER:  33,595
;      REFERENCE/DOCKET NUMBER:  GTR90-04 PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312-856-7180
;      TELEFAX:  312-856-4972
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1196 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
PCT-US92-06617A-7


  Query Match             13.4%;  Score 26;  DB 5;  Length 1196;
  Best Local Similarity   46.9%;  Pred. No. 5;
  Matches   69;  Conservative    3;  Mismatches   75;  Indels    0;  Gaps    0;

Qy       1 ATGGTTACAAGTGTAAGGACTCAGCCCCCCGTCATAATGCCAGGTATGCAGACCGAGATC 60
           | |||  ||  | |  | |||:   :  |      ||  |   |  | ||||| ||| |
Db     634 AAGGTGGCACATATCCGCACTYGCNHAACATATGGAACCCGGCGGCTCCAGACGGAGCTG 575

Qy      61 AAAACGCAGGCCACGAATCTTGCGGCGAATCTTTCCGCAGTCAGAGAAAGTGCCACAGCG 120
             |  | | | |  |   |||  | ||     |     |||   | |||   |  ||:
Db     574 GCAGAGAATGGCATCATCGTTGGCGTGACCGACTGGCACGTCTTCGTAAGGAGCTAAGST 515

Qy     121 ACGCTGTCAGGGGAAATTAAAGGCCCG 147
           ||||||| ||  ||||   |||  | |
Db     514 ACGCTGTAAGCAGAAACGCAAGTTCAG 488


RESULT  13
US-08-390-878-17
; Sequence 17, Application US/08390878
; Patent No. 5700683
;  GENERAL INFORMATION:
;    APPLICANT:  Stover, Charles K.
;    APPLICANT:  Mahairas, Gregory G.
;    TITLE OF INVENTION:  VIRULENCE-ATTENUATING GENETIC DELETIONS
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend Khourie and Crew
;      STREET:  One Market Plaza, Steuart Street Tower, 20th
;      STREET:  Floor
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/390,878
;      FILING DATE:  17-FEB-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hunter, Tom
;      REGISTRATION NUMBER:  38,498
;      REFERENCE/DOCKET NUMBER:  15371A-17
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/543/9600
;      TELEFAX:  415/543/5043
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  15239 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-390-878-17


  Query Match             13.3%;  Score 25.8;  DB 2;  Length 15239;
  Best Local Similarity   54.8%;  Pred. No. 15;
```

```
  Matches   60;  Conservative    0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy      19 ACTCAGCCCCCCGTCATAATGCCAGGTATGCAGACCGAGATCAAAACGCAGGCCACGAAT 78
           ||| | ||  ||   | | | | | |   | |||  || |   ||| ||  |  |  ||
Db     415 ACTGATCCTGCCAGAAAATTTCAAAGACAGTAGATTGAAAAGGAAAGGCTTGGTAAAAAA 474

Qy      79 CTTGCGGCGAATCTTTCCGCAGTCAGAGAAAGTGCCACAGCGACGCTGTCAGGGGA 134
             | | |  | || |  |  |||  || | ||||   |||     ||| |||| ||
Db     475 GGTTCAGGCATTCCTAGCCGAGTGTGACACAGTGGAGCAGAACATCTGCCAGGAGA 530


RESULT   9
US-08-461-359-3
; Sequence 3, Application US/08461359
; Patent No. 5686595
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, John C.
;    APPLICANT:  Takayama, Shinichi
;    APPLICANT:  Sato, Takaaki
;    TITLE OF INVENTION:  Bcl-2-associated proteins
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CAMPBELL AND FLORES
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/461,359
;      FILING DATE:  10-NOV-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-LJ 9725
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  733 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-461-359-3


  Query Match             13.6%;  Score 26.4;  DB 2;  Length 733;
  Best Local Similarity   51.7%;  Pred. No. 3.1;
  Matches   60;  Conservative    0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy      19 ACTCAGCCCCCCGTCATAATGCCAGGTATGCAGACCGAGATCAAAACGCAGGCCACGAAT 78
           ||| | ||  ||   | | | | | |   | |||  || |   ||| ||  |  |  ||
Db     415 ACTGATCCTGCCAGAAAATTTCAAAGACAGTAGATTGAAAAGGAAAGGCTTGGTAAAAAA 474

Qy      79 CTTGCGGCGAATCTTTCCGCAGTCAGAGAAAGTGCCACAGCGACGCTGTCAGGGGA 134
             | | |  | || |  |  |||  || | ||||   |||     ||| |||| ||
Db     475 GGTTCAGGCATTCCTAGCCGAGTGTGACACAGTGGAGCAGAACATCTGCCAGGAGA 530


RESULT  10
PCT-US94-12904-3
; Sequence 3, Application PC/TUS9412904
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, John C.
;    APPLICANT:  Takayama, Shinichi
;    APPLICANT:  Sato, Takaaki
;    TITLE OF INVENTION:  Bcl-2-associated proteins
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CAMPBELL AND FLORES
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/12904
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/152,485
;      FILING DATE:  12-NOV-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  FP-LJ 1201
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  733 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US94-12904-3


  Query Match             13.6%;  Score 26.4;  DB 5;  Length 733;
  Best Local Similarity   51.7%;  Pred. No. 3.1;
  Matches   60;  Conservative    0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy      19 ACTCAGCCCCCCGTCATAATGCCAGGTATGCAGACCGAGATCAAAACGCAGGCCACGAAT 78
           ||| | ||  ||   | | | | | |   | |||  || |   ||| ||  |  |  ||
Db     415 ACTGATCCTGCCAGAAAATTTCAAAGACAGTAGATTGAAAAGGAAAGGCTTGGTAAAAAA 474

Qy      79 CTTGCGGCGAATCTTTCCGCAGTCAGAGAAAGTGCCACAGCGACGCTGTCAGGGGA 134
             | | |  | || |  |  |||  || | ||||   |||     ||| |||| ||
Db     475 GGTTCAGGCATTCCTAGCCGAGTGTGACACAGTGGAGCAGAACATCTGCCAGGAGA 530


RESULT  11
US-08-375-241-7/c
; Sequence 7, Application US/08375241
; Patent No. 5648481
;  GENERAL INFORMATION:
;    APPLICANT:  Parodos, Kyriaki
;    APPLICANT:  McCarty, Janice
;    TITLE OF INVENTION:  Nucleic Acid Probes for the Detection of
;    TITLE OF INVENTION:  Shigella
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
;      STREET:  Two Militia Drive
;      CITY:  Lexington
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
```

```
;       ADDRESSEE:  CAMPBELL AND FLORES
;       STREET:  4370 La Jolla Village Drive, Suite 700
;       CITY:  San Diego
;       STATE:  California
;       COUNTRY:  USA
;       ZIP:  92122
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/152,485
;       FILING DATE:  10-NOV-1993
;       CLASSIFICATION:  436
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Campbell, Cathryn A.
;       REGISTRATION NUMBER:  31,815
;       REFERENCE/DOCKET NUMBER:  P-LJ 9725
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (619) 535-9001
;       TELEFAX:  (619) 535-8949
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  733 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-08-152-485-3


  Query Match             13.6%;  Score 26.4;  DB 1;  Length 733;
  Best Local Similarity   51.7%;  Pred. No. 3.1;
  Matches   60;  Conservative    0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy      19 ACTCAGCCCCCCGTCATAATGCCAGGTATGCAGACCGAGATCAAAACGCAGGCCACGAAT 78
           ||| | ||  ||   | | | | | |   | |||  || |  ||| ||  |  | ||
Db     415 ACTGATCCTGCCAGAAAATTTCAAAGACAGTAGATTGAAAAGGAAAGGCTTGGTAAAAAA 474

Qy      79 CTTGCGGCGAATCTTTCCGCAGTCAGAGAAAGTGCCACAGCGACGCTGTCAGGGGA 134
             | | |  | || |  |  |||  || | ||||   |||     ||| |||| ||
Db     475 GGTTCAGGCATTCCTAGCCGAGTGTGACACAGTGGAGCAGAACATCTGCCAGGAGA 530


RESULT  7
US-08-463-089-3
; Sequence 3, Application US/08463089
; Patent No. 5641866
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, John C.
;    APPLICANT:  Takayama, Shinichi
;    APPLICANT:  Sato, Takaaki
;    TITLE OF INVENTION:  Bcl-2-associated proteins
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CAMPBELL AND FLORES
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/463,089
;      FILING DATE:  10-NOV-1993
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-LJ 9725
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  733 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-463-089-3


  Query Match             13.6%;  Score 26.4;  DB 1;  Length 733;
  Best Local Similarity   51.7%;  Pred. No. 3.1;
  Matches   60;  Conservative    0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy      19 ACTCAGCCCCCCGTCATAATGCCAGGTATGCAGACCGAGATCAAAACGCAGGCCACGAAT 78
           ||| | ||  ||   | | | | | |   | |||  || |  ||| ||  |  | ||
Db     415 ACTGATCCTGCCAGAAAATTTCAAAGACAGTAGATTGAAAAGGAAAGGCTTGGTAAAAAA 474

Qy      79 CTTGCGGCGAATCTTTCCGCAGTCAGAGAAAGTGCCACAGCGACGCTGTCAGGGGA 134
             | | |  | || |  |  |||  || | ||||   |||     ||| |||| ||
Db     475 GGTTCAGGCATTCCTAGCCGAGTGTGACACAGTGGAGCAGAACATCTGCCAGGAGA 530


RESULT  8
US-08-461-360A-3
; Sequence 3, Application US/08461360A
; Patent No. 5650491
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, John C.
;    APPLICANT:  Takayama, Shinichi
;    APPLICANT:  Sato, Takaaki
;    TITLE OF INVENTION:  Bcl-2-associated proteins
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CAMPBELL AND FLORES
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/461,360A
;      FILING DATE:  10-NOV-1993
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-LJ 9725
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  733 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-461-360A-3


  Query Match             13.6%;  Score 26.4;  DB 1;  Length 733;
  Best Local Similarity   51.7%;  Pred. No. 3.1;
```

```
US-08-486-269A-3
; Sequence 3, Application US/08486269A
; Patent No. 5945509
;  GENERAL INFORMATION:
;    APPLICANT:  Heinemann, Stephen F.
;    APPLICANT:  Boulter, James R.
;    APPLICANT:  Hollmann, Michael
;    APPLICANT:  Bettler, Bernhard
;    APPLICANT:  Jensen, Jan E.
;    TITLE OF INVENTION:  GLUTAMATE RECEPTOR COMPOSITIONS
;    TITLE OF INVENTION:  AND METHODS
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Gray Cary Ware & Freidenrich LLP
;      STREET:  4365 Executive Drive, Suite 1600
;      CITY:  San Diego
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92121
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows DEMONSTRATION Version 2.0D
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/486,269A
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/013,767
;      FILING DATE:  04-FEB-1993
;      APPLICATION NUMBER:  07/718,575
;      FILING DATE:  21-JUN-1991
;      APPLICATION NUMBER:  PCT/US90/06153
;      FILING DATE:  25-OCT-1990
;      APPLICATION NUMBER:  07/428,116
;      FILING DATE:  27-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Reiter, Stephen E
;      REGISTRATION NUMBER:  31,192
;      REFERENCE/DOCKET NUMBER:  P41 9986
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-677-1409
;      TELEFAX:  619-677-1465
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3505 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    IMMEDIATE SOURCE:
;      CLONE:  GluR2
;    FEATURE:
;      NAME/KEY:  Coding Sequence
;      LOCATION:  316...2964
;      OTHER INFORMATION:
US-08-486-269A-3
```

```
  Query Match             14.4%;  Score 28;  DB 4;  Length 3505;
  Best Local Similarity   48.2%;  Pred. No. 1.6;
  Matches   79;  Conservative    0;  Mismatches   85;  Indels    0;  Gaps    0;

Qy      10 AGTGTAAGGACTCAGCCCCCCGTCATAATGCCAGGTATGCAGACCGAGATCAAAACGCAG 69
           ||  ||   || || |  ||   || ||| ||  | ||   |       || | | | |
Db    2873 AGAATATTAACCCATCTTCCTCGCAGAATTCCCAGAATTTTGCAACTTATAAGGAAGGTT 2932

Qy      70 GCCACGAATCTTGCGGCGAATCTTTCCGCAGTCAGAGAAAGTGCCACAGCGACGCTGTCA 129
           | ||| || | ||  |||   | |    | | || | | |   |   || || |   | |
Db    2933 ACAACGTATATGGCATCGAGAGTGTTAAAATTTAGGGGATGACCTTGAGTGATGTCATGA 2992
```

```
Qy     130 GGGGAAATTAAAGGCCCGCAACTGGAAGATTTTCCCGCGCTGAT 173
           || | ||   |||||   ||| |  | ||  | |  | |   ||
Db    2993 GGAGCAAGGCAAGGCTGTCAATTACAGGAAGTACTGGAGAAAAT 3036
```

```
RESULT   5
PCT-US93-03076-1
; Sequence 1, Application PC/TUS9303076
;  GENERAL INFORMATION:
;    APPLICANT:  Whitehead Institute for Biomedical Research
;    TITLE OF INVENTION:  GAP-Associated Protein p190 and
;    TITLE OF INVENTION:  Transduction
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
;      STREET:  2 Militia Drive
;      CITY:  Lexington
;      STATE:  MA
;      COUNTRY:  US
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/03076
;      FILING DATE:  19930331
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Granahan, Patricia
;      REGISTRATION NUMBER:  32,227
;      REFERENCE/DOCKET NUMBER:  WHI92-03A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-861-6240
;      TELEFAX:  617-861-9540
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8298 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  731..5272
PCT-US93-03076-1
```

```
  Query Match             14.0%;  Score 27.2;  DB 5;  Length 8298;
  Best Local Similarity   58.8%;  Pred. No. 4;
  Matches   47;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy       3 GGTTACAAGTGTAAGGACTCAGCCCCCCGTCATAATGCCAGGTATGCAGACCGAGATCAA 62
           ||  |||||| |  ||||| || |  | | |||  || |||   |||||  |  | |||
Db    2851 GGACACAAGTGGAGAGACTCTGCACAGCTTAATACAGCAAGGCCAGCAGATTGCTAGCAA 2910

Qy      63 AACGCAGGCCACGAATCTTG 82
           |   |||        |||||
Db    2911 ACTTCAGTGTGTCTTTCTTG 2930
```

```
RESULT   6
US-08-152-485-3
; Sequence 3, Application US/08152485
; Patent No. 5539094
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, John C.
;    APPLICANT:  Takayama, Shinichi
;    APPLICANT:  Sato, Takaaki
;    TITLE OF INVENTION:  Bcl-2-associated proteins
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
```

```
          ||||||   |  ||||
Db   3339 AAACAGCATAAGCTGG 3324


RESULT   2
US-07-718-575-3
; Sequence 3, Application US/07718575
; Patent No. 5202257
;  GENERAL INFORMATION:
;    APPLICANT:  Heinemann Ph.D., Stephen F.
;    APPLICANT:  Boulter Ph.D., James R.
;    APPLICANT:  Hollmann Ph.D., Michael NMN
;    APPLICANT:  Bettler Ph.D., Bernhard NMN
;    APPLICANT:  Jensen Ph.D., Jan E.
;    TITLE OF INVENTION:  GLUTAMATE RECEPTOR COMPOSITIONS AND
;    TITLE OF INVENTION:  METHODS
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pretty, Schroeder, Brueggemann & Clark
;      STREET:  444 So. Flower St., Suite 2000
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  90071-2921
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/718,575
;      FILING DATE:  19910813
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Reiter Ph.D., Stephen E.
;      REGISTRATION NUMBER:  31192
;      REFERENCE/DOCKET NUMBER:  P31 8962
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619)535-8949
;      TELEX:  9103330318
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3505 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    IMMEDIATE SOURCE:
;      CLONE:  GluR2
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  316..2967
US-07-718-575-3


  Query Match             14.4%;  Score 28;  DB 1;  Length 3505;
  Best Local Similarity   48.2%;  Pred. No. 1.6;
  Matches   79;  Conservative    0;  Mismatches   85;  Indels    0;  Gaps    0;

Qy      10 AGTGTAAGGACTCAGCCCCCCGTCATAATGCCAGGTATGCAGACCGAGATCAAAACGCAG 69
           ||  ||   || || |  ||   || ||| ||  | ||   |      || |  | |
Db    2873 AGAATATTAACCCATCTTCCTCGCAGAATTTCCCAGAATTTTGCAACTTATAAGGAAGGTT 2932

Qy      70 GCCACGAATCTTGCGGCGAATCTTTCCGCAGTCAGAGAAAGTGCCACAGCGACGCTGTCA 129
           | ||| || | ||  |||   | |    | | || | | |  |   || || |   | |
Db    2933 ACAACGTATATGGCATCGAGAGTGTTAAAATTTAGGGGATGACCTTGAGTGATGTCATGA 2992

Qy     130 GGGGAAATTAAAGGCCCGCAACTGGAAGATTTTCCCGCGCTGAT 173
           || | ||   |||||   ||| |  | ||  | |  | |   ||
Db    2993 GGAGCAAGGCAAGGCTGTCAATTACAGGAAGTACTGGAGAAAAT 3036
```

```
RESULT   3
US-08-481-206-3
; Sequence 3, Application US/08481206
; Patent No. 5739291
;  GENERAL INFORMATION:
;    APPLICANT:  Heinemann Ph.D., Stephen F.
;    APPLICANT:  Boulter Ph.D., James R.
;    APPLICANT:  Hollmann Ph.D., Michael NMN
;    APPLICANT:  Bettler Ph.D., Bernhard NMN
;    APPLICANT:  Jensen Ph.D., Jan E.
;    TITLE OF INVENTION:  GLUTAMATE RECEPTOR COMPOSITIONS AND
;    TITLE OF INVENTION:  METHODS
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pretty, Schroeder, Brueggemann & Clark
;      STREET:  444 So. Flower St., Suite 2000
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  90071-2921
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/481,206
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/013,767
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Reiter Ph.D., Stephen E.
;      REGISTRATION NUMBER:  31192
;      REFERENCE/DOCKET NUMBER:  P31 8962
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619)535-8949
;      TELEX:  9103330318
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3505 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    IMMEDIATE SOURCE:
;      CLONE:  GluR2
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  316..2967
US-08-481-206-3


  Query Match             14.4%;  Score 28;  DB 2;  Length 3505;
  Best Local Similarity   48.2%;  Pred. No. 1.6;
  Matches   79;  Conservative    0;  Mismatches   85;  Indels    0;  Gaps    0;

Qy      10 AGTGTAAGGACTCAGCCCCCCGTCATAATGCCAGGTATGCAGACCGAGATCAAAACGCAG 69
           ||  ||   || || |  ||   || ||| ||  | ||   |      || |  | |
Db    2873 AGAATATTAACCCATCTTCCTCGCAGAATTCCCAGAATTTTGCAACTTATAAGGAAGGTT 2932

Qy      70 GCCACGAATCTTGCGGCGAATCTTTCCGCAGTCAGAGAAAGTGCCACAGCGACGCTGTCA 129
           | ||| || | ||  |||   | |    | | || | | |  |   || || |   | |
Db    2933 ACAACGTATATGGCATCGAGAGTGTTAAAATTTAGGGGATGACCTTGAGTGATGTCATGA 2992

Qy     130 GGGGAAATTAAAGGCCCGCAACTGGAAGATTTTCCCGCGCTGAT 173
           || | ||   |||||   ||| |  | ||  | |  | |   ||
Db    2993 GGAGCAAGGCAAGGCTGTCAATTACAGGAAGTACTGGAGAAAAT 3036


RESULT   4
```

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 1999, 18:01:01 ; Search time 140.41 Seconds
(without alignments)
143.753 Million cell updates/sec

Title: US-09-068-804-4
Perfect score: 194
Sequence: 1 ATGGTTACAAGTGTAAGGAC.....AAACAGGCGAGTCTGGATGC 194

Scoring table: IDENTITY_NUC

Searched: 192659 seqs, 52021692 residues

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/5C_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/5D_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	30.4	15.7	4060	1	US-08-308-949A-1	Sequence 1, Appli
	2	28	14.4	3505	1	US-07-718-575-3	Sequence 3, Appli
	3	28	14.4	3505	2	US-08-481-206-3	Sequence 3, Appli
	4	28	14.4	3505	4	US-08-486-269A-3	Sequence 3, Appli
	5	27.2	14.0	8298	5	PCT-US93-03076-1	Sequence 1, Appli
	6	26.4	13.6	733	1	US-08-152-485-3	Sequence 3, Appli
	7	26.4	13.6	733	1	US-08-463-089-3	Sequence 3, Appli
	8	26.4	13.6	733	1	US-08-461-360A-3	Sequence 3, Appli
	9	26.4	13.6	733	2	US-08-461-359-3	Sequence 3, Appli
	10	26.4	13.6	733	5	PCT-US94-12904-3	Sequence 3, Appli
c	11	26	13.4	1196	1	US-08-375-241-7	Sequence 7, Appli
c	12	26	13.4	1196	5	PCT-US92-06617A-7	Sequence 7, Appli
	13	25.8	13.3	15239	2	US-08-390-878-17	Sequence 17, Appl
c	14	25.6	13.2	9126	3	US-08-580-038-26	Sequence 26, Appl
c	15	25.6	13.2	9126	4	US-08-639-857-3	Sequence 3, Appli
c	16	25.4	13.1	1174	1	US-08-375-241-5	Sequence 5, Appli
	17	25.4	13.1	4040	3	US-08-685-118-1	Sequence 1, Appli
	18	25.4	13.1	4040	3	US-08-915-495-1	Sequence 1, Appli
	19	25.4	13.1	4040	4	US-08-914-520-1	Sequence 1, Appli
c	20	25.4	13.1	1174	5	PCT-US92-06617A-5	Sequence 5, Appli
	21	25	12.9	4852	1	US-07-853-913-3	Sequence 3, Appli
	22	24.8	12.8	1834	2	US-08-297-633A-1	Sequence 1, Appli
	23	24.8	12.8	1448	2	US-08-585-758A-1	Sequence 1, Appli
	24	24.8	12.8	1448	3	US-08-977-818-1	Sequence 1, Appli
	25	24.8	12.8	1834	3	US-08-485-721-8	Sequence 8, Appli
	26	24.8	12.8	4434	3	US-08-540-406-5	Sequence 5, Appli
	27	24.8	12.8	1834	3	US-08-392-935-8	Sequence 8, Appli
	28	24.8	12.8	3900	3	US-08-356-060A-42	Sequence 42, Appl
	29	24.8	12.8	1448	4	US-08-670-274B-1	Sequence 1, Appli
c	30	24.8	12.8	5467	4	US-08-605-106-7	Sequence 7, Appli
	31	24.8	12.8	1834	5	PCT-US93-08325-1	Sequence 1, Appli
	32	24.8	12.8	1834	5	PCT-US93-08326-8	Sequence 8, Appli
	33	24.8	12.8	4434	5	PCT-US95-13233-5	Sequence 5, Appli
	34	24.6	12.7	1534	4	US-08-850-880-1	Sequence 1, Appli
	35	24.6	12.7	1534	4	US-08-944-916-1	Sequence 1, Appli
c	36	24.4	12.6	5399	2	US-08-368-071-9	Sequence 9, Appli
c	37	24.4	12.6	5399	3	US-08-458-181-9	Sequence 9, Appli
c	38	24.4	12.6	5399	5	PCT-US93-02172-9	Sequence 9, Appli
c	39	24.2	12.5	396	1	US-08-470-179-40	Sequence 40, Appl
c	40	24.2	12.5	3531	3	US-08-629-600-1	Sequence 1, Appli
	41	24	12.4	2375	2	US-08-369-796-9	Sequence 9, Appli
	42	24	12.4	2375	4	US-08-852-091-9	Sequence 9, Appli
	43	24	12.4	2375	5	PCT-US95-17025-9	Sequence 9, Appli
c	44	23.8	12.3	15664	1	US-08-402-282-3	Sequence 3, Appli
	45	23.8	12.3	410	6	5432081-1	Patent No. 5432081

ALIGNMENTS

RESULT 1
US-08-308-949A-1/c
; Sequence 1, Application US/08308949A
; Patent No. 5580703
; GENERAL INFORMATION:
; APPLICANT: Kotin, Robert M.
; APPLICANT: Berns, Kenneth I.
; APPLICANT: Linden, Ralph M.
; TITLE OF INVENTION: Human Adeno-Associated Virus Integration
; TITLE OF INVENTION: Site DNA and Uses Thereof
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/308,949A
; FILING DATE: September 20, 1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/947,127
; FILING DATE: September 27, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Carroll, Alice O.
; REGISTRATION NUMBER: 33,542
; REFERENCE/DOCKET NUMBER: ACC92-10F
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4060 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-308-949A-1

Query Match 15.7%; Score 30.4; DB 1; Length 4060;
Best Local Similarity 51.5%; Pred. No. 0.25;
Matches 70; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

```
Qy      55 GAGATCAAAACGCAGGCCACGAATCTTGCGGCGAATCTTTCCGCAGTCAGAGAAAGTGCC 114
           || |   | | ||||    ||| || || | |||||     ||  |||||| |||| |
Db    3459 GAAACACAGCCCCAGGTGGAGAAACTGGCCGGGAATCAAGAGTCACCCAGAGACAGTGAC 3400

Qy     115 ACAGCGACGCTGTCAGGGGAAATTAAAGGCCCGCAACTGGAAGATTTTCCCGCGCTGATC 174
             | |  | ||||      |      |||    || |   || | |   | |  |||  |
Db    3399 CAACCATCCCTGTTTTCCTAGGACTGAGGGTTTCAGCGCTAAAACTAGGCTGTCCTGGGC 3340

Qy     175 AAACAGGCGAGTCTGG 190
```

This Page Blank (uspto)

```
Db   1138 ACCGACATGAATCAGGATTTGAATGCTCTGGCAAATAATGTCACGACTAAAGCGAATGAG 1197

Qy    301 GTTGTGCAAACCCAGTTACGCGAGCAGCAGGCAGAAGTCGGAAAGTTTTTTGATATTAGC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1198 GTTGTGCAAACCCAGTTACGCGAGCAGCAGGCAGAAGTCGGAAAGTTTTTTGATATTAGC 1257

Qy    361 GGAATGTCTTCCAGTGCCGTTGCGCTGTTGGCTGCCGCGAATACGTTAATGCTGACGTTG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1258 GGAATGTCTTCCAGTGCCGTTGCGCTGTTGGCTGCCGCGAATACGTTAATGCTGACGTTG 1317

Qy    421 AACCAGGCTGATAGCAAACTGTCTGGTAAGTTGTCATTAGTCAGTTTTGATGCAGCTAAA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1318 AACCAGGCTGATAGCAAACTGTCTGGTAAGTTGTCATTAGTCAGTTTTGATGCAGCTAAA 1377

Qy    481 ACGACGGCAAGCTCCATGATGCGCGAAGGGATGAATGCGTTGTCCGGTAGTATTTCCCAG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1378 ACGACGGCAAGCTCCATGATGCGCGAAGGGATGAATGCGTTGTCCGGTAGTATTTCCCAG 1437

Qy    541 AGCGCGCTTCAGTTGGGGATCACTGGCGTGGGCGCCAAACTGGAATATAAGGGGCTGCAG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1438 AGCGCGCTTCAGTTGGGGATCACTGGCGTGGGCGCCAAACTGGAATATAAGGGGCTGCAG 1497

Qy    601 AATGAAAGAGGCGCGCTTAAACATAATG-CCGCGAAGATCGATAAACTGACCACTGAAAG 659
          |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db   1498 AATGAAAGAGGCGCGCTTAAACATAATGCCCGCGAAGATCGATAAACTGACCACTGAAAG 1557

Qy    660 CCACAGTATTAAAAACGTGCTGAACGGGCAGAATAGCGTCAAACTCGGTGCTGAAGGCGT 719
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   1558 CCACAGTATTAAAAACGTGC-GAACGGGCAGAATAGCGTCAAACTCGGTGCTGAAGGCGT 1616

Qy    720 CGATTCTCTGAAATCGTTAAATATGAAGAAAACCGGTACCGATGCGACGAAAAATCTTAA 779
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1617 CGATTCTCTGAAATCGTTAAATATGAAGAAAACCGGTACCGATGCGACGAAAAATCTTAA 1676

Qy    780 TGATGCGACGCTTAAATCTAATGCCGGAACCAGCGCCACGGAAAGTCTGGGTATTAAAGA 839
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1677 TGATGCGACGCTTAAATCTAATGCCGGAACCAGCGCCACGGAAAGTCTGGGTATTAAAGA 1736

Qy    840 CAGTAATAAACAAATCTCCCCTGAACATCAGGCTATTCTGTCGAAACGTCTTGAGTCTGT 899
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1737 CAGTAATAAACAAATCTCCCCTGAACATCAGGCTATTCTGTCGAAACGTCTTGAGTCTGT 1796

Qy    900 CGAATCCGATATTCGTCTTGAGCAGAATACCATGGATATGACCCGAATCGATGCGCGCAA 959
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1797 CGAATCCGATATTCGTCTTGAGCAGAATACCATGGATATGACCCGAATCGATGCGCGCAA 1856

Qy    960 GATGCAGATGACGGGCGATCTGATTATGAAGAACTCGGTCACGGTCGGTGGTATTGCAGG 1019
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1857 GATGCAGATGACGGGCGATCTGATTATGAAGAACTCGGTCACGGTCGGTGGTATTGCAGG 1916

Qy   1020 GGCGTCCGGGCAGTACGCCGCTACTCAGGAACGTTCCGAGCAGCAAATTAGCCAGGTGAA 1079
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1917 GGCGTCCGGGCAGTACGCCGCTACTCAGGAACGTTCCGAGCAGCAAATTAGCCAGGTGAA 1976

Qy   1080 TAACCGGGTTGCCAGCACCGCATCGGACGAAGCCCGTGAAAGTTCACGTAAATCGACCAG 1139
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1977 TAACCGGGTTGCCAGCACCGCATCGGACGAAGCCCGTGAAAGTTCACGTAAATCGACCAG 2036

Qy   1140 CCTGATTCAGGAAATGCTGAAAACAATGGAGAGCATTAACCAGTCGAAAGCATCCGCACT 1199
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2037 CCTGATTCAGGAAATGCTGAAAACAATGGAGAGCATTAACCAGTCGAAAGCATCCGCACT 2096

Qy   1200 CGCTGCTATCGCAGGCAATATTCGCGCTTAA 1230
          ||||||||||||||||||||||||||||||
Db   2097 CGCTGCTATCGCAGGCAATATTCGCGCTTAA 2127


RESULT   3
Q91286
ID   Q91286 standard; DNA; 4940 BP.
AC   Q91286;
DT   25-NOV-1995  (first entry)
DE   ORC1 gene encoding a subunit of yeast origin of DNA replication.
KW   Origin of replication complex; ARS element; subunit; ss.
OS   Saccharomyces cerevisiae.
FH   Key             Location/Qualifiers
FT   cds             682. .3423
FT                   /*tag= a
PN   WO9516694-A.
PD   22-JUN-1995.
PF   16-DEC-1994; U14563.
PR   16-DEC-1993; US-168479.
PA   (COLD-) COLD SPRING HARBOR LAB.
PA   (REGC ) UNIV CALIFORNIA.
PI   Bell SP,  Foss M,  Herskowitz I,  Kobayashi R,  Laurenson P;
PI   Li JJ,  Mcnally FJ,  Rine J,  Stillman BW;
DR   WPI; 95-231507/30.
DR   P-PSDB; R77274.
PT   New origin of replication complex genes and proteins - used to
PT   identify agents useful in the diagnosis, prognosis and treatment
PT   of disease
PS   Claim 2; Page 42-44; 70pp; English.
CC   A subset of yeast ARS elements act as chromosome origins of DNA
CC   replication. A multi-protein complex that recognises the ARS
CC   consensus sequence (ACS) is called the origin recognition complex
CC   (ORC). It bind to more than 10 yeast ARS elements. It mediates the
CC   function of ACS. To clone the genes encoding the various ORC
CC   subunits, ORC was extracted from BJ296 cells and ORC subunits were
CC   digested with Achromobacter protease I. The purified peptides were
CC   sequenced. To clone the gene for the largest (120kD) subunit of ORC,
CC   degenerate oligo primers 1201 and 1202 were synthesised based on the
CC   sequence of the first ORC1 peptide. These oligos were used to perform
CC   PCR using total yeast genomic DNA from strain W303 as target. A
CC   48 bp fragment was amplified, subcloned and sequenced. A
CC   radiolabeled probe of the PCR product was then used to probe a
CC   genomic library of yeast DNA sequences resulting in the
CC   identification of two overlapping clones. Sequencing of these
CC   clones resulted in the identification of a large ORF that encoded
CC   a protein with a predicted mol. wt. of 120 kd and that encoded all
CC   four ORC1 peptide sequences.
SQ   Sequence   4940 BP;    1671 A;    849 C;    993 G;    1427 T;


  Query Match             2.8%;  Score 35;  DB 1;  Length 4940;
  Best Local Similarity   54.2%;  Pred. No. 0.47;
  Matches   71;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy    739 AATATGAAGAAAACCGGTACCGATGCGACGAAAAATCTTAATGATGCGACGCTTAAATCT 798
          ||   ||  ||||  ||| |  |   ||   | ||   || | |  || | |  |  ||
Db   1334 AAAGAGATCAAAAGAGGTCCTCAAAAGAAAGATAAGGCTACTCAAACGGCACAAATTTCA 1393

Qy    799 AATGCCGGAACCAGCGCCACGGAAAGTCTGGGTATTAAAGACAGTAATAAACAAATCTCC 858
          | || | ||| || || || || |    || || | | ||| ||||| || |    ||
Db   1394 GACGCAGAAACAAGAGCTACAGATATAACGGATAATGAGGACGGTAATGAAGATGAATCA 1453

Qy    859 CCTGAACATCA 869
           ||||  || |
Db   1454 TCTGATTATGA 1464


RESULT   4
T73279
ID   T73279 standard; cDNA; 4940 BP.
AC   T73279;
DT   12-SEP-1997  (first entry)
DE   S. cerevisiae origin of replication complex protein 1 gene.
KW   Origin of replication complex; ORC; yeast; Saccharomyces cerevisiae;
KW   chromatography; peptide sequencing; primer; amplification; PCR; genome;
KW   polymerase chain reaction; open reading frame; cell growth; cancer;
KW   infection; inflammation; hypersensitivity; ds.
OS   Saccharomyces cerevisiae.
FH   Key             Location/Qualifiers
FT   cds             683. .3427
FT                   /*tag= a
FT                   /product= ORC1 protein
```

```
PN   US5614618-A.
PD   25-MAR-1997.
PF   16-DEC-1993; 168479.
PR   16-DEC-1993; US-168479.
PR   07-JUN-1995; US-484106.
PA   (COLD-) COLD SPRING HARBOR LAB.
PA   (REGC ) UNIV CALIFORNIA.
PI   Bell SP,  Foss M,  Gavin K,  Herskowitz I,  Hidaka M;
PI   Kobayashi R,  Laurenson P,  Li J,  McNally FJ,  Rine J;
PI   Stillman BW;
DR   WPI; 97-201534/18.
DR   P-PSDB; W22224.
PT   Nucleic acids encoding origin of replication complex proteins - used
PT   for screening for lead cpds. for therapy or diagnosis of disease
PT   associated with undesirable cell growth
PS   Claim 4; Column 19-24; 53pp; English.
CC   This is the nucleotide sequence encoding the origin of replication
CC   complex protein 1 (ORC1) from the yeast Saccharomyces cerevisiae.
CC   The ORC proteins were isolated by conventional chromatographic methods
CC   and peptide sequence of the proteins determined.  Primers were then
CC   synthesised based on the peptide sequence from ORC1 and used to amplify
CC   a 48 bp fragment.  The fragment was then used to screen a yeast genomic
CC   library.  Clones isolated from the library were sequenced and used to
CC   construct the complete ORC1 coding sequence.  The open reading frame
CC   encodes a protein of 914 amino acid with a molecular weight 120 kD.
CC   The ORC proteins (W22224-35) can be used to screen chemical libraries
CC   to identify lead compounds useful in treatment and diagnosis of
CC   undesired cell growth, e.g. cancer, infections, inflammation and
CC   hypersensitivity.
SQ   Sequence   4940 BP;    1671 A;    849 C;    993 G;    1427 T;


  Query Match              2.8%;  Score 35;  DB 1;  Length 4940;
  Best Local Similarity   54.2%;  Pred. No. 0.47;
  Matches   71;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy     739 AATATGAAGAAAACCGGTACCGATGCGACGAAAAATCTTAATGATGCGACGCTTAAATCT 798
           ||   ||  ||||  ||| |  |   ||   | ||   || | |  || | |  |  ||
Db    1334 AAAGAGATCAAAAGAGGTCCTCAAAAGAAAGATAAGGCTACTCAAACGGCACAAATTTCA 1393

Qy     799 AATGCCGGAACCAGCGCCACGGAAAGTCTGGGTATTAAAGACAGTAATAAACAAATCTCC 858
           | || | ||| || || || || |    || || | | ||| ||||| || |    ||
Db    1394 GACGCAGAAACAAGAGCTACAGATATAACGGATAATGAGGACGGTAATGAAGATGAATCA 1453

Qy     859 CCTGAACATCA 869
           ||||  || |
Db    1454 TCTGATTATGA 1464


RESULT   5
T41476
ID   T41476 standard; DNA; 7291 BP.
AC   T41476;
DT   15-JAN-1997  (first entry)
DE   Haemophilus adhesion protein HA2 gene.
KW   Haemophilus adhesion protein; HA2; hsf protein; vaccine; ss.
OS   Haemophilus influenzae type b strain C54.
FH   Key             Location/Qualifiers
FT   misc_difference 4
FT                   /*tag=  a
FT                   /note= "base 4 is given as n in the specification"
FT   misc_difference 57
FT                   /*tag=  b
FT                   /note= "base 57 is given as n in the specification"
FT   misc_difference 59
FT                   /*tag=  c
FT                   /note= "base 59 is given as n in the specification"
FT   cds             163. .7224
FT                   /*tag=  d
PN   WO9630519-A1.
PD   03-OCT-1996.
PF   22-MAR-1996; U04031.
PR   24-MAR-1995; US-409995.
PA   (UYSL-) UNIV ST LOUIS.
PA   (UNIW ) UNIV WASHINGTON.
PI   Barenkamp SJ,  St Geme JW;
DR   WPI; 96-455364/45.
DR   P-PSDB; R99393.
PT   Recombinant Haemophilus adhesion proteins HA1 and HA2 - for use in
PT   vaccines against H. influenzae infection.
PS   Claim 8; Page 55-65; 120pp; English.
CC   A genomic DNA clone (T41476) codes for the Haemophilus adhesion
CC   protein HA2 (R99393) of H. influenzae type b strain C54.  It was
CC   isolated from a 8.3 kb XbaI fragment of plasmid pDC400, a pUC19
CC   deirvative contg. the C54 surface fibril locus.  HA2 is associated
CC   with the formation of surface fibrils involved in adhesion to host
CC   cells.  The isolated DNA can be incorporated into a vector and used
CC   to direct expression of recombinant HA2 in transformed host cells
CC   for use in vaccines against H. influenzae infection.
SQ   Sequence   7291 BP;    2508 A;    1316 C;    1758 G;    1706 T;


  Query Match              2.8%;  Score 34.4;  DB 1;  Length 7291;
  Best Local Similarity   47.3%;  Pred. No. 0.91;
  Matches  104;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy     619 AAACATAATGCCGCGAAGATCGATAAACTGACCACTGAAAGCCACAGTATTAAAAACGTG 678
           ||||||| ||   |||     | |  |    |||  |   |  |  || || |||| |
Db    3970 AAACATACTGTAACGATTGATGTTGCAGAAGCCAAAGTTGGTGATGGTCTTGAAAAAGAT 4029

Qy     679 CTGAACGGGCAGAATAGCGTCAAACTCGGTGCTGAAGGCGTCGATTCTCTGAAATCGTTA 738
               ||||  ||| ||   ||||| | | |  |  ||  |   ||  |||   | | |
Db    4030 ACTGACGGCAAGATTAAACTCAAAGTAGATAATACAGATGGGAATAATCTATTAACCGTT 4089

Qy     739 AATATGAAGAAAACCGGTACCGATGCGACGAAAAATCTTAATGATGCGACGCTTAAATCT 798
           ||   |  |||   |   ||| ||| | |    |  ||||||  |  ||    |  |
Db    4090 GATGCAACAAAAGGTGCATCCGTTGCCAAGGGCGAGTTTAATGCCGTAACAACAGATGCA 4149

Qy     799 AATGCCGGAACCAGCGCCACGGAAAGTCTGGGTATTAAAG 838
           | | | |       || | |  | | |  | |  ||||
Db    4150 ACTACAGCCCAAGGCACAAATGCCAATGAGCGCGGTAAAG 4189


RESULT   6
Q63296
ID   Q63296 standard; DNA; 459 BP.
AC   Q63296;
DT   05-JAN-1995  (first entry)
DE   Consensus sequence of mannuronan C-5-epimerase gene R blocks.
KW   Mannuronan; epimerase; alginates; immunostimulant;
KW   cell immobilisation; Azotobacter vinelandii; guluronic acid;
KW   mannuronic acid; ss.
OS   Synthetic.
PN   WO9409124-A.
PD   28-APR-1994.
PF   08-OCT-1993; NO0151.
PR   08-OCT-1992; GB-021163.
PA   (NOBI-) NOBIPOL NOBIPOLS FORSKNINGSSTIFTELSE.
PA   (PRON-) PRONOVA BIOPOLYMER AS.
PI   Ertesvag H,  Larsen B,  Valla S,  Skjak-braek G;
DR   WPI; 94-151310/18.
PT   New nucleic acid encoding mannuronan C-5-epimerase - and derived
PT   vectors and polypeptide, for prodn. of alginate of controlled G
PT   or M content, useful e.g. as immunostimulants or for cell
PT   immobilisation.
PS   Disclosure; Page 51-52; 78pp; English.
CC   The epimerase enzymes (See R54628-R54631) can be used for the
CC   microbial production of alginates, esp. those of G (guluronic acid)
CC   block content 75-98% or those with a specific M (mannuronic acid)/G
CC   block content.  Inactivation of specific enzyme activities allows
CC   prodn. of pure poly-M alginates or those of G block content below
CC   25%.  Alginates of high M content are powerful, nontoxic
CC   immunostimulants while those of high G content can be used for cell
CC   immobilisation.  This sequence is the consensus of the epimerase
CC   enzyme gene R blocks.  R blocks are designated regions found within
CC   the genes encoding the epimerase enzymes.
```

```
SQ   Sequence   459 BP;    80 A;    185 C;    148 G;    46 T;

  Query Match             2.8%;  Score 34;  DB 1;  Length 459;
  Best Local Similarity   50.6%;  Pred. No. 0.28;
  Matches   82;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy     552 GTTGGGGATCACTGGCGTGGGCGCCAAACTGGAATATAAGGGGCTGCAGAATGAAAGAGG 611
           | ||||  |||| ||  |||||| |       |    |    ||||| | |    |  |
Db     273 GCTGGGCTTCACCGGGCTGGGCGACGGCTACAACGGCACCCTGCTGCTGCAGCTCAACGC 332

Qy     612 CGCGCTTAAACATAATGCCGCGAAGATCGATAAACTGACCACTGAAAGCCACAGTATTAA 671
           |||    |  |  |    |  ||||| | |  |   |  | | ||  ||||  |  |  |
Db     333 CGCCGGCACCCGCACCTACCTGAAGAGCTACGAGGCGGACGCCGAGGGCCAGCGCTTCGA 392

Qy     672 AAACGTGCTGAACGGGCAGAATAGCGTCAAACTCGGTGCTGA 713
           | ||  ||| ||||  |  | | || | | ||||| || ||
Db     393 GATCGCCCTGGACGGCAACTACACCGGCCAGCTCGGCGCCGA 434


RESULT   7
X53491/c
ID   X53491 standard; DNA; 114955 BP.
AC   X53491;
DT   05-JUL-1999  (first entry)
DE   Human adenosine A1 receptor antisense oligonucleotide fragment.
KW   Antisense oligonucleotide; multiple target; antisense treatment;
KW   impaired respiration; inflammation; lung disease;
KW   pulmonary vasoconstriction; inflammation; allergic rhinitis;
KW   acute asthma; allergy; asthma; impeded respiration;
KW   respiratory distress syndrome; pain; cystic fibrosis;
KW   pulmonary hypertension; pulmonary vasoconstriction; emphysema;
KW   chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;
KW   colon cancer; breast cancer; lung cancer; pancreatic cancer;
KW   hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;
KW   prostate cancer; ss.
OS   Synthetic.
PN   WO9913886-A1.
PD   25-MAR-1999.
PF   17-SEP-1998; U19419.
PR   09-JUN-1998; US-093972.
PR   17-SEP-1997; US-059160.
PA   (UYEC-) UNIV EAST CAROLINA.
PI   Nyce JW;
DR   WPI; 99-229400/19.
PT   New antisense oligonucleotides used in treatment of, e.g. pulmonary
PT   vasoconstriction
PS   Disclosure; Page 37; 120pp; English.
CC   The specification describes antisense oligonucleotides (X52869-X55271)
CC   directed against at least 2 mRNAs selected from target genes, coding and
CC   non-coding regions of RNAs corresponding to target genes, gene
CC   initiation codons, genomic flanking regions, intron-exon borders, the
CC   5'-end, the 3'-end and the juxta-section between coding and non-coding
CC   regions and all segments of RNAs encoding proteins associated with one
CC   or more diseases, conditions or mixtures. The antisense oligonucleotides
CC   may be derived from sequences X55272-74. These multiple target
CC   oligonucleotides (specifically X55180-271) can be used for the antisense
CC   treatment of diseases and conditions. Typical diseases and conditions
CC   are those associated with impaired respiration and inflammation,
CC   including lung diseases, pulmonary vasoconstriction, inflammation,
CC   allergic rhinitis, acute asthma, allergies, asthma, impeded respiration,
CC   respiratory distress syndrome, pain, cystic fibrosis, pulmonary
CC   hypertension, pulmonary vasoconstriction, emphysema, chronic obstructive
CC   pulmonary disease (COPD), and cancers such as leukemias, lymphomas,
CC   carcinomas e.g. colon cancer, breast cancer, lung cancer, pancreatic
CC   cancer, hepatocellular carcinoma, kidney cancer, melanoma, hepatic
CC   metastases, as well as all types of cancers which may metastasize or have
CC   metastasized to the lungs, including breast and prostate cancer.
SQ   Sequence   114955 BP;    6071 A;    29417 C;    36712 G;    21328 T;

  Query Match             2.7%;  Score 33.8;  DB 1;  Length 114955;
  Best Local Similarity   45.5%;  Pred. No. 6.2;
  Matches   76;  Conservative    8;  Mismatches   83;  Indels    0;  Gaps    0;

Qy     550 CAGTTGGGGATCACTGGCGTGGGCGCCAAACTGGAATATAAGGGGCTGCAGAATGAAAGA 609
           | |  | | |  || |||  |||| |||| | ||||    |||  |     |     | |
Db   53342 CGGAGGAGCAGGACCGGCAAGGGCACCAAGCAGGAACGAGAGGAACGAAGCAGGAGGACA 53283

Qy     610 GGCGCGCTTAAACATAATGCCGCGAAGATCGATAAACTGACCACTGAAAGCCACAGTATT 669
           | | |      :    |   |:|| ::| :|| |||  |  ||| |||||: |:|| |
Db   53282 GACACAGGCNNHNNNSACAGCVCGGVVAAVGAAAAAAVGGACACAGAAAGVAAVAGGAGA 53223

Qy     670 AAAAACGTGCTGAACGGGCAGAATAGCGTCAAACTCGGTGCTGAAGG 716
             |||:      || |  |  ||  |  ||| | |   | ||| |
Db   53222 GCAAAVNNHNNNSACAGCTCGGTTAATGAAAAAATGGACACAGAAAG 53176


RESULT   8
V74898
ID   V74898 standard; DNA; 4709 BP.
AC   V74898;
DT   16-MAR-1999  (first entry)
DE   Staphylococcus aureus contig SEQ ID #587.
KW   Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW   cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW   skin infection; surgical wound infection; scalded skin syndrome;
KW   toxic shock syndrome; ds.
OS   Staphylococcus aureus.
FH   Key             Location/Qualifiers
FT   misc_feature    1681. .1740
FT                   /*tag= a
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
FT                          are included to maintain the nucleotide numbering
FT                          given in the specification for this DNA sequence"
FT   misc_feature    3481. .3540
FT                   /*tag= b
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
FT                          are included to maintain the nucleotide numbering
FT                          given in the specification for this DNA sequence"
PN   EP-786519-A2.
PD   30-JUL-1997.
PF   07-JAN-1997; 100117.
PR   05-JAN-1996; US-009861.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI   Rosen CA;
DR   WPI; 97-374922/35.
PT   Polynucleotide(s) and proteins derived from Staphylococcus aureus -
PT   stored on computer readable medium and used in the production of
PT   anti-S.aureus vaccines
PS   Claim 1; Page 1519-1521; 3271pp; English.
CC   This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC   of the invention. The DNA sequences are recorded on a computer readable
CC   medium, preferably selected from a floppy or hard disk, random access
CC   memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC   the S.aureus DNA sequences allows putative functions to be assigned so
CC   that protein-encoding or regulatory regions of commercial, therapeutic or
CC   industrial importance can be obtained. Specifically, sequences which are
CC   likely to encode antigens have been identified and these polypeptides can
CC   be used in a vaccine composition against S.aureus infection. The
CC   polypeptides can also be used in a kit for the immunodetection of
CC   S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC   including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC   skin and surgical wound infections, scalded skin syndrome, toxic shock
CC   syndrome, etc. Organisms transformed with the DNA sequences can be used
CC   for recombinant production of the polypeptides. The new DNA sequences
CC   (and their fragments) are useful as primers or probes for isolating
CC   homologues of any of the S.aureus DNA sequences contained on the
CC   computer readable medium.
SQ   Sequence   4709 BP;    1657 A;    704 C;    908 G;    1313 T;

  Query Match             2.7%;  Score 33.4;  DB 1;  Length 4709;
```

```
  Best Local Similarity   36.6%;  Pred. No. 1.5;
  Matches   64;  Conservative    0;  Mismatches  111;  Indels    0;  Gaps    0;

Qy     601 AATGAAAGAGGCGCGCTTAAACATAATGCCGCGAAGATCGATAAACTGACCACTGAAAGC 660
           || ||| || || | || ||| ||  || |  ||||    |  |  ||||||
Db    1607 AAAGAAGCAGATGCAAGTGAAAATAGTGTTACGCAATCTGATAGCGCAAGTAACGAAAGC 1666

Qy     661 CACAGTATTAAAAACGTGCTGAACGGGCAGAATAGCGTCAAACTCGGTGCTGAAGGCGTC 720
           | |||| | |
Db    1667 AAAAGTAATGATTCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1726

Qy     721 GATTCTCTGAAATCGTTAAATATGAAGAAAACCGGTACCGATGCGACGAAAAATC 775
                       || ||| |  || ||  ||  ||  || | || ||||||
Db    1727 NNNNNNNNNNNNNGTCAAACACTAATAATGGCGAAACGAGTGTGGCGCAAAATC 1781


RESULT   9
V73327
ID   V73327 standard; DNA; 1282 BP.
AC   V73327;
DT   25-FEB-1999  (first entry)
DE   Human CD18 pseudogene clone 1 DNA.
KW   Pseudogene; CD18; human; cytokeratin 18; detection; epithelial cell;
KW   metastases; tumour; ss.
OS   Homo sapiens.
PN   DE19716346-C1.
PD   19-NOV-1998.
PF   18-APR-1997; 016346.
PR   18-APR-1997; DE-016346.
PA   (WAGE/) WAGENER C.
PI   Neumaier M, Tschentscher P, Wagener C;
DR   WPI; 98-584374/50.
PT   Oligonucleotide primers for amplifying cytokeratin 18 cDNA  -
PT   especially for detecting metastatic epithelial tumour cells
PS   Claim 5; Fig 1; 10pp; German.
CC   This sequence is the human cytokeratin 18 (CK18) pseudogene clone 1 which
CC   is used in a method to produce oligonucleotide primers capable of
CC   detecting the CK18 pseudogene. The oligonucleotides are useful as primers
CC   in an assay for detecting the CK18 gene in a sample, especially an organ
CC   (e.g. bone-marrow) smear, puncture or biopsy sample or a blood, sputum,
CC   urine, stool, liquor, bile, lymph or gastrointestinal secretion sample.
CC   The assay can be used to detect epithelial cells, especially metastases
CC   of tumour cells of epithelial origin. The primers have higher affinity
CC   for CK18 cDNA than for processed CK18 pseudogenes. The detection limit
CC   of the assay is 1-10 epithelial cells per ml blood.
SQ   Sequence   1282 BP;    345 A;    333 C;    369 G;    235 T;


  Query Match             2.7%;  Score 33.2;  DB 1;  Length 1282;
  Best Local Similarity   48.0%;  Pred. No. 0.88;
  Matches   95;  Conservative    0;  Mismatches  103;  Indels    0;  Gaps    0;

Qy     644 AACTGACCACTGAAAGCCACAGTATTAAAAACGTGCTGAACGGGCAGAATAGCGTCAAAC 703
           | |||   || || | | |   | | |||||   ||||  |     ||| | |    ||
Db     589 AGCTGGAGACAGAGATCGAGGCTCTCAAAAAGCAGCTGCTCTTCATGAAGAACCATGAAA 648

Qy     704 TCGGTGCTGAAGGCGTCGATTCTCTGAAATCGTTAAATATGAAGAAAACCGGTACCGATG 763
             |  |   ||||| |  |  ||| ||   |       |  |  ||  |  |  || ||||
Db     649 AGGAAGTAAAAGGCCTACAAGCTCAGATTGCCAGCTCTGGGTTGACCATGGAGACAGATG 708

Qy     764 CGACGAAAAATCTTAATGATGCGACGCTTAAATCTAATGCCGGAACCAGCGCCACGGAAA 823
           |  | |||  ||   |  |||| | | | |   |  |   | | | ||   | |||
Db     709 CCCCCAAATCTCAGGACCATGCCAAGATCATGGCAGACATCAGGAACAATACGACGAGCT 768

Qy     824 GTCTGGGTATTAAAGACA 841
           | ||||| |  |  || |
Db     769 GGCTGGGAAGAACTGAGA 786


RESULT  10
Q23079
ID   Q23079 standard; DNA; 513 BP.
AC   Q23079;
DT   17-AUG-1992  (first entry)
DE   Antigen mc-30c gene.
KW   Oocysts; sporozoite; beta-galactosidase; rat antiserum; ss.
OS   Eimeria maxima.
FH   Key             Location/Qualifiers
FT   tata_signal     158. .163
FT                   /*tag= a
FT                   /number= 1
FT   tata_signal     172. .175
FT                   /*tag= b
FT                   /number= 2
FT   cds             1. .513
FT                   /*tag= c
FT                   /product= mc-30c
FT   cds             184. .513
FT                   /*tag= d
FT                   /note= "Eimeria antigen (109 amino acids, 12.2 kD)"
PN   WO9204461-A.
PD   19-MAR-1992.
PF   05-SEP-1991; U06431.
PR   12-SEP-1990; US-581694.
PA   (GENE-) GENEX CORP.
PA   (FARH ) HOECHST AG.
PI   Jacobson JW, Strausberg RL, Wilson SD, Pope SH, Strausberg SL;
PI   Raether W;
DR   WPI; 92-114366/14.
DR   P-PSDB; R22379.
PT   Vaccine against avian coccidiosis - comprising recombinant
PT   Eimeria antigen mc-4c, mc-5c or mc-30c gene, etc., or
PT   microorganisms expressing them
PS   Claim 5; Page 64 + Fig 5; 94pp; English.
CC   To identify antigens of E. maxima, expression libraries were prepd.
CC   in lambda vector, lambda gt11, using cDNA prepd. from polyA mRNA
CC   isolated from E. maxima oocysts.  The cDNA expression library was
CC   screened with rat antiserum raised against the sporozoite stage of
CC   E. maxima.  The library to be screened was plated on a host that
CC   allows lysis and plaque formation.  Following induction of the
CC   antigens encoded by the phage, the plaques were transferred to
CC   nitrocellulose filters.  Positive phage were identified after
CC   screening with the rat anti-E. maxima sporozoite antiserum.
CC   The cDNA inserts from the positive clones were cloned into
CC   bacteriophage M13 and subjected to sequence analysis.
CC   E. maxima antigen mc-30c was identified.
CC   Antigen mc-30c has a 510 bp ORF fused in frame with the beta-
CC   galactosidase gene of lambdagt11.  Within the ORF are two
CC   methionine codons (bases 184-189) followed by 321 bp of ORF ending
CC   with a termination codon.  Preceding the putative methionine
CC   initiation codons are two TATA boxes.  Thus it is possible that the
CC   insert contains a complete gene encoding an Eimeria antigen
CC   consisting of 109 amino acids with a mol. wt. of 12.2 kD.
SQ   Sequence   513 BP;    198 A;    55 C;    113 G;    147 T;


  Query Match             2.7%;  Score 32.8;  DB 1;  Length 513;
  Best Local Similarity   56.5%;  Pred. No. 0.73;
  Matches   61;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy     245 ACATGAATCAGGATTTGAATGCTCTGGCAAATAATGTCACGACTAAAGCGAATGAGGTTG 304
           | ||  ||||| ||  ||    || | |||  |||| |  |   | |  | | ||| |||
Db      12 AGATACATCAGCATGGGATACATCCGTCAAGGAATGGCTTGTTGATACAGGAAGAGTTTG 71

Qy     305 TGCAAACCCAGTTACGCGAGCAGCAGGCAGAAGTCGGAAAGTTTTTTG 352
           ||||    ||||  |  |   |||||    | |||| | | ||  |||
Db      72 TGCAGGTGCAGTAGCAAGCTTAGCAGATGAAGGTCGTATATTTGGTTG 119


RESULT  11
N60392
ID   N60392 standard; DNA; 1612 BP.
AC   N60392;
DT   27-JUN-1991  (first entry)
DE   Sequence encoding the Asparagine-Rich Protein (ARP)
DE   Ag319 of Plasmodium falciparum.
```

```
KW   Malaria vaccine; ss.
OS   Plasmodium falciparum.
FH   Key             Location/Qualifiers
FT   cds             2. .1612
FT                   /*tag= a
PN   WO8606075-A.
PD   23-OCT-1986.
PF   11-APR-1986; AU0092.
PR   11-APR-1985; AU-001008.
PR   11-APR-1985; AU-000108.
PR   25-JUL-1985; AU-001640.
PR   11-APR-1986; AU-056037.
PA   (HALL-) HALL W & E INST MED.
PA   (KEMP/) KEMP D J.
PI   Kemp D, Anders R, Brown G, Coppel RL;
DR   WPI; 86-291646/44.
DR   P-PSDB; P60452.
PT   New DNA molecules encoding Plasmodium falciparum antigens -
PT   namely the SHARP ARP and MESA antigens
PS   Claim 6; Fig 8; 52pp; English.
CC   The antigens of the invention are prod. using recombinant DNA
CC   techniques. They may be used in vaccine compsns. to stimulate an
CC   immune response against P.falciparum, the cause of human malaria.
SQ   Sequence   1612 BP;    831 A;    148 C;    178 G;    455 T;
```

```
  Query Match              2.6%;  Score 31.6;  DB 1;  Length 1612;
  Best Local Similarity   46.1%;  Pred. No. 3.3;
  Matches  106;  Conservative    0;  Mismatches  124;  Indels    0;  Gaps    0;

Qy     619 AAACATAATGCCGCGAAGATCGATAAACTGACCACTGAAAGCCACAGTATTAAAAACGTG 678
           ||  ||| |       || ||  ||||| |||    | |    || | ||  |||||
Db     740 AATAATATTAATAATAATATGAATAAAATGAATTATAATTATCAAAATAGAAAAAATAGT 799

Qy     679 CTGAACGGGCAGAATAGCGTCAAACTCGGTGCTGAAGGCGTCGATTCTCTGAAATCGTTA 738
           ||||     | |  | |   ||  |   |  | |      | ||    ||||
Db     800 ATGAATAATAATATGAACAATAATATGAATAATAATATGAACCATAACATGAATAATAAT 859

Qy     739 AATATGAAGAAAACCGGTACCGATGCGACGAAAAATCTTAATGATGCGACGCTTAAATCT 798
           ||||||||  | | |   | | ||   | ||| ||| | |||      | | | || | |
Db     860 AATATGAACCATAACATGAACAATAACATGAACAATATAAATATGAGTATGATGAATTAT 919

Qy     799 AATGCCGGAACCAGCGCCACGGAAAGTCTGGGTATTAAAGACAGTAATAA 848
           |||              ||| |  | |  ||  | |||  | |  |||||
Db     920 AATAATAATGAATCTAACACAGCTAATAGGGAAAATAACAATATGAATAA 969
```

```
RESULT  12
Q35072
ID   Q35072 standard; DNA; 565 BP.
AC   Q35072;
DT   20-MAY-1993  (first entry)
DE   HCV envelope region nucleic acid.
KW   Envelope; region; type C; hepatitis; virus; HCV; vaccine; serum;
KW   non-A, non-B; amplify; ss.
OS   Hepatitis C virus.
PN   J04349885-A.
PD   04-DEC-1992.
PF   29-MAY-1991; 152169.
PR   29-MAY-1991; JP-152169.
PA   (TEIJ ) TEIJIN LTD.
DR   WPI; 93-022708/03.
PT   Envelope region nucleic acid fragment - for type C hepatitis
PT   virus (I), for producing vaccine
PS   Claim 1; Page 2; 13pp; Japanese.
CC   This sequence encodes a novel envelope region of type C hepatitis
CC   virus (HCV).  This fragment can be used for the preparation of a
CC   vaccine for hepatitis C.  This fragment was prepared from the serum
CC   of non-A, non-B hepatitis patients and the envelope region DNA was
CC   amplified by PCR using the primer sequences given in Q35073-76.
SQ   Sequence   565 BP;    61 A;    92 C;    106 G;    85 T;
```

```
  Query Match              2.6%;  Score 31.6;  DB 1;  Length 565;
  Best Local Similarity   26.8%;  Pred. No. 1.9;
  Matches   91;  Conservative   89;  Mismatches  158;  Indels    2;  Gaps    1;

Qy     622 CATAATGCCGCGAAGATCGATAAACTGACCACTGAAAGCCACAGTATTAAAAACGTGCTG 681
           |: || |:: |   | :  |: |  | ||:| :||  |:: :|     |  |  |||:
Db     139 CRCAACGSSWCMGGGNYRTAYCATGTCACBARYGAYTGYYCYAACKCRAGYATWGTGYAY 198

Qy     682 AACGGGCAGAATAGCGTCAAACTCGGTGCTGAAGGCGTCGATT--CTCTGAAATCGTTAA 739
           :| | : :| |:: : ||:  |::: : |:|   |:::: :|   |: |  |  :: :
Db     199 RARGYRRMGGAYRYYATCMTGCWYDYYCCBGGGTGYRYBCCYTGYGTYCGSGARRRYRRY 258

Qy     740 ATATGAAGAAAACCGGTACCGATGCGACGAAAAATCTTAATGATGCGACGCTTAAATCTA 799
           |  :|      ||:: || |:   |  |   |:   :  | :   :::   : :|
Db     259 NNYTCBMGNTGYTGGGYRGCGCTBACYCCCACGSTYGCYRCYAGRRAYVBYAVYVTYCCC 318

Qy     800 ATGCCGGAACCAGCGCCACGGAAAGTCTGGGTATTAAAGACAGTAATAAACAAATCTCCC 859
           : : :|:::: :  :|  |  :  | :: | |  |:  |:::|  :    |  || |:
Db     319 RCBRYGVMRMTWCGBCGBCAYRTCGAYYTGCTYGTHGGGRSVGCYRCYYTCTGYTCSGCY 378

Qy     860 CTGAACATCAGGCTATTCTGTCGAAACGTCTTGAGTCTGTCGAATCCGATATTCGTCTTG 919
           :|: |: |  |:   :|:||  |:  :||:||   : | ||  |:: | |   | |||
Db     379 VTSTAYGTKGGRGAYYTBTGCGGRTCYGTYTTYCTYRTYTCYCABYTGTTTACCYTCTMB 438

Qy     920 AGCAGAATACCATGGATATGACCCGAATCGATGCGCGCAA 959
           : ::| :: :   :|| :: :|  |: |  |   |: |::
Db     439 MCYMGSMDGYAYGDGARRDYRCARGRYTGYAAYTGYTCDM 478
```

```
RESULT  13
X14251
ID   X14251 standard; DNA; 1363 BP.
AC   X14251;
DT   31-MAR-1999  (first entry)
DE   H. pylori GHPO 1449 gene.
KW   GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW   peptic ulcer disease; ss.
OS   Helicobacter pylori.
FH   Key             Location/Qualifiers
FT   CDS             51. .1313
FT                   /*tag= a
PN   WO9843478-A1.
PD   08-OCT-1998.
PF   01-APR-1998; U06371.
PR   29-JUL-1997; US-902615.
PR   01-APR-1997; US-833457.
PR   24-JUN-1997; US-881227.
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (INMR ) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
PI   Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;
DR   WPI; 98-542293/46.
DR   P-PSDB; W98532.
PT   New isolated Helicobacter polynucleotides - used to develop products
PT   for the diagnosis, prevention and treatment of Helicobacter
PT   infections and gastrointestinal diseases
PS   Claim 1; Page 1016-1018; 2054pp; English.
CC   This sequence represents a polynucleotide of the invention. It was
CC   isolated from Helicobacter pylori and encodes a H.pylori GHPO protein.
CC   The polypeptides can be used for preventing or treating Helicobacter
CC   infections, and gastroduodenal diseases associated with these
CC   infections, including acute, chronic, and atrophic gastritis, and peptic
CC   ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used
CC   for the production of antibodies. The products can also be used for
CC   detection and diagnosis.
SQ   Sequence   1363 BP;    439 A;    224 C;    313 G;    387 T;
```

```
  Query Match              2.6%;  Score 31.6;  DB 1;  Length 1363;
  Best Local Similarity   56.9%;  Pred. No. 3;
  Matches   58;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy     665 GTATTAAAAACGTGCTGAACGGGCAGAATAGCGTCAAACTCGGTGCTGAAGGCGTCGATT 724
           |||| || | | |  || | ||| ||    ||||  ||  ||  |  ||| || ||| ||
```

```
Db    547 GTATCAACACCTTAATGCAAGGGGAGGGCTGCGTTCAAAGCGTGGTGGAAAGCATCGTTT 606

Qy    725 CTCTGAAATCGTTAAATATGAAGAAAACCGGTACCGATGCGA 766
Db    607 ATGCGGATAGTTTTCATGACACAAAAAACGCTTTTGATGCGA 648


RESULT  14
Q66074/c
ID   Q66074 standard; DNA; 2973 BP.
AC   Q66074;
DT   09-NOV-1994  (first entry)
DE   Sequence of microsomal delta-12 desaturase gene.
KW   Fatty acid; desaturase; lipid; unsaturated; transgenic plant; ss.
OS   Arabidopsis thaliana.
FH   Key             Location/Qualifiers
FT   exon            433. .520
FT                   /*tag=  a
FT   intron          521. .1654
FT                   /*tag=  b
PN   WO9411516-A.
PD   26-MAY-1994.
PF   15-OCT-1993; U09987.
PR   17-NOV-1992; US-977339.
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
PI   Lightner JE,  Okuley JJ;
DR   WPI; 94-183515/22.
PT   Genes for fatty acid desaturase enzymes - permit alteration of
PT   plant lipid composition
PS   Claim 3; Page 136-138; 147pp; English.
CC   The gene corresp. to Q66068 was isolated by screening Arabidopsis
CC   genomic DNA library using radiolabeled pSF2b cDNA insert, purifying
CC   positively-hybridising plaque, and subcloning a 6kb Hind III insert
CC   fragment from the phage DNA in pBluescript vector. Comparison of the
CC   sequences of the gene (Q66074) and the cDNA (Q66068) revealed the
CC   presence of a single intron of 1134 bp at a posn. between
CC   nucleotides 88 and 89 or the cDNA, which is 4 nucleotides 5' to the
CC   initiation codon.  The cDNA is contained in clone p92103. Plasmid
CC   p92103 was deposited on October 16, 1992 with the ATCC and bears
CC   accession number ATCC 69095. An isolated nucleic acid fragment
CC   where in the nucleic acid identity is 90% or greater to
CC   Q66068 or Q66074 is claimed.
SQ   Sequence   2973 BP;    762 A;    637 C;    569 G;    1005 T;


  Query Match             2.6%;  Score 31.4;  DB 1;  Length 2973;
  Best Local Similarity   57.7%;  Pred. No. 5.3;
  Matches   56;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy    657 AAGCCACAGTATTAAAAACGTGCTGAACGGGCAGAATAGCGTCAAACTCGGTGCTGAAGG 716
Db    567 AAGAATCAAAATCGAAAATGAGAAGAAATGGAAGAGAAGCGACGGACCTGGAGATGGAGC 508

Qy    717 CGTCGATTCTCTGAAATCGTTAAATATGAAGAAAACC 753
Db    507 TGACGTAGGGGCGATAACGTTAATAACGATGAACACC 471


RESULT  15
T75616
ID   T75616 standard; DNA; 10592 BP.
AC   T75616;
DT   20-FEB-1998  (first entry)
DE   Canine herpesvirus genomic DNA nCUS10592.
KW   CHV; vaccine; vector; gene therapy; canid; dog; CIR6; CUS2;
KW   CPK; serine-threonine protein kinase US3; CgD; glycoprotein D;
KW   CgG; glycoprotein G; glycoprotein I; CgI; CUS85; CUS9;
KW   unique short region; ds.
OS   Canine herpesvirus.
FH   Key             Location/Qualifiers
FT   CDS             complement (377. .921)
FT                   /*tag=  a
FT                   /product= CIR6
FT                   /note= "CHV IR6 (CIR6) gene"
FT   CDS             complement (1080. .2254)
FT                   /*tag=  b
FT                   /product= CUS2
FT                   /note= "CHV US2 (CUS2) gene"
FT   CDS             2375. .3577
FT                   /*tag=  c
FT                   /product= CPK
FT                   /note= "CHV serine-threonine protein kinase US3
FT                         (CPK) gene"
FT   CDS             3689. .4936
FT                   /*tag=  d
FT                   /product= CgG
FT                   /note= "CHV glycoprotein G (CgG) gene"
FT   CDS             5128. .6165
FT                   /*tag=  e
FT                   /product= CgD
FT                   /note= "CHV glycoprotein D (CgD) gene"
FT   CDS             6225. .7319
FT                   /*tag=  f
FT                   /product= CgI
FT                   /note= "CHV glycoprotein I (CgI) gene"
FT   CDS             7469. .9035
FT                   /*tag=  g
FT                   /product= CgE
FT                   /note= "CHV glycoprotein E (CgE) gene"
FT   CDS             9028. .9264
FT                   /*tag=  h
FT                   /product= CUS8.5
FT                   /note= "CHV unique short region 8.5 gene"
FT   CDS             9376. .9735
FT                   /*tag=  i
FT                   /product= CUS9
FT                   /note= "CHV unique short region 9 gene"
PN   WO9729772-A1.
PD   21-AUG-1997.
PF   14-FEB-1997; U04115.
PR   15-FEB-1996; US-602010.
PA   (HESK-) HESKA CORP.
PI   Frank RA,  Haanes EJ;
DR   WPI; 97-424758/39.
DR   P-PSDB; W22979-83 AND W22998-3002.
PT   Recombinant canine herpes virus and its genome - useful as vaccine
PT   to protect canids against infectious, metabolic or genetic diseases
PS   Claim 22; Page 164-170; 240pp; English.
CC   This DNA sequence comprises genomic DNA molecule nCUS10592 of
CC   canine herpesvirus (CHV).  It includes coding sequences for CIR6,
CC   CUS2, CPK and CgG proteins (see W22979-82) and for CgD, CgI, CgE,
CC   CUS8.5 and CUS9 proteins (see W22998-3002).  CHV nucleic acids,
CC   encoded proteins and antibodies can be used to protect canids
CC   against CHV infection.  Live recombinant CHV can be used as a
CC   vaccine vector to protect canids against infectious, metabolic or
CC   genetic diseases, providing a vigourous and long-lasting immunity.
CC   The genome can accommodate several sections of heterologous DNA
CC   (for making a multivalent vaccine) without interfering with
CC   packaging to live virus.  Fragments of CHV nucleic acids may also
CC   be used as probes and primers, and to inhibit CHV infection.
SQ   Sequence   10592 BP;    3757 A;    1536 C;    1681 G;    3618 T;


  Query Match             2.6%;  Score 31.4;  DB 1;  Length 10592;
  Best Local Similarity   52.7%;  Pred. No. 10;
  Matches   68;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy      6 AATTAGTAATGTGGGAATAAATCCCGCCGCTTATTTAAATAATCATTCTGTTGAGAATAG 65
Db   9016 AATCAATCCAGAATGAATAAATCTAAACTCTCATTTAAAGAAAAAAACGCTATATATGAA 9075

Qy     66 TTCACAGACAGCTTCGCAATCCGTTAGCGCTAAAGATATTCTGAATAGTATTGGTATTAG 125
Db   9076 TTTAAAAATATTTTATCAAACACTTCATTGTCAACTTTTCCTGTATTATCGTTTAATGAG 9135

Qy    126 CAGCAGTAA 134
```

```
              ||| ||
Db    9136 GAGCCAAAA 9144
```

Search completed: October 5, 1999, 17:58:46
Job time: 11368 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 1999, 17:55:59 ; Search time 2414.69 Seconds
(without alignments)
1619.993 Million cell updates/sec

Title: US-09-068-804-2
Perfect score: 1230
Sequence: 1 ATGTTAATTAGTAATGTGGG.....CAGGCAATATTCGCGCTTAA 1230

Scoring table: IDENTITY_NUC

Searched: 679419 seqs, 1590154680 residues

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_st:*
14: gb_sts:*
15: gb_sy:*
16: gb_un:*
17: gb_vi:*
18: em_fun:*
19: em_htg:*
20: em_hum1:*
21: em_hum2:*
22: em_in:*
23: em_om:*
24: em_or:*
25: em_ov:*
26: em_pat:*
27: em_ph:*
28: em_pl:*
29: em_ro:*
30: em_sts:*
31: em_sy:*
32: em_un:*
33: em_vi:*
34: gb_htg1:*
35: gb_htg2:*
36: gb_in1:*
37: gb_in2:*
38: em_ba1:*
39: em_ba2:*
40: em_hum3:*
41: em_hum4:*
42: gb_pr4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1230	100.0	3441	2	STU30491	U30491 Salmonella
2	1222	99.3	5393	1	STSIPGNS	X82670 S.typhi spa
3	1157	94.1	3622	1	STU25631	U25631 Salmonella
4	551.2	44.8	2995	1	SDU66877	U66877 Salmonella
5	66.4	5.4	5255	1	SHFIPABCD	J04117 S.flexneri
6	66.4	5.4	4761	1	W10PPIBCA	M34849 Plasmid pWR
7	64.8	5.3	7663	1	SFPVR2	X15319 Shigella fl
8	63.2	5.1	6869	1	SDIPABCD	X60777 Shigella dy
c 9	50.8	4.1	7218	5	I66494	I66494 Sequence 14
c 10	41	3.3	103694	42	AC004836	AC004836 Homo sapi
11	41	3.3	204597	42	AC005069	AC005069 Homo sapi
12	40.2	3.3	119660	35	AC006424	AC006424 Arabidops
13	38.8	3.2	153773	11	HUAC004020	AC004020 Homo sapi
14	37.8	3.1	1080	10	HSDGCR6	X96484 H.sapiens m
15	37.2	3.0	462	36	LM5UTRDGN	Y09088 L.major mRN
c 16	36.8	3.0	88413	8	ATAC005724	AC005724 Arabidops
c 17	36	2.9	936	36	MSSSURRNA	Z69606 M.secans SS
18	35.6	2.9	13939	36	LMA237587	AJ237587 Leishmani
19	35.2	2.9	76458	8	F28K20	AC004793 Arabidops
c 20	35	2.8	338534	1	ECOUW93	U14003 Escherichia
c 21	35	2.8	11313	2	AE000502	AE000502 Escherich
22	35	2.8	4940	5	I32730	I32730 Sequence 1
23	35	2.8	4940	5	I38706	I38706 Sequence 1
24	35	2.8	19975	7	SC9745	Z38114 S.cerevisia
25	35	2.8	5166	8	SCU34860	U34860 Saccharomyc
c 26	34.8	2.8	1644	9	HSA9668	AJ009668 Homo sapi
c 27	34.8	2.8	1707	11	HSU26742	U26742 Human dystr
c 28	34.8	2.8	351	12	MMMTHFR05	AF105992 Mus muscu
29	34.6	2.8	75188	7	AB005238	AB005238 Arabidops
30	34.4	2.8	7062	2	U41852	U41852 Haemophilus
31	34.4	2.8	5738	5	I55123	I55123 Sequence 3
32	34.4	2.8	130926	9	HS453A3	Z83745 Human DNA s
33	34.4	2.8	168753	11	AC006120	AC006120 Homo sapi
34	34.4	2.8	300172	35	AC005308	AC005308 Plasmodiu
c 35	34.2	2.8	188477	34	AC004971	AC004971 Homo sapi
36	34	2.8	212150	1	BSUB0020	Z99123 Bacillus su
c 37	34	2.8	65143	1	D83026	D83026 Bacillus su
c 38	34	2.8	37310	36	CEZK1098	Z22176 Caenorhabdi
c 39	34	2.8	111448	37	AC005334	AC005334 Drosophil
40	33.8	2.7	126150	11	AC005152	AC005152 Homo sapi
41	33.8	2.7	348	11	AF052785	AF052785 Homo sapi
42	33.8	2.7	258033	35	AC005507	AC005507 Plasmodiu
c 43	33.6	2.7	10109	2	MPAE000048	AE000048 Mycoplasm
c 44	33.6	2.7	2781	9	HSU09609	U09609 Human p80HT
c 45	33.4	2.7	5249	1	ECFIMCLUS	Z37500 E.coli type

ALIGNMENTS

RESULT 1
STU30491
LOCUS STU30491 3441 bp DNA BCT 19-JAN-1996
DEFINITION Salmonella typhimurium SspA (sspA) and SspB (sspB) genes, partial cds, and SspC (sspC) and SspD (sspD) genes, complete cds.
ACCESSION U30491
NID g1117811
VERSION U30491.1 GI:1117811
KEYWORDS .
SOURCE Salmonella typhimurium.
ORGANISM Salmonella typhimurium
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Salmonella.
REFERENCE 1 (bases 1 to 3441)
AUTHORS Hueck,C.J., Hantman,M.J., Bajaj,V., Johnston,C., Lee,C.A. and Miller,S.I.
TITLE Salmonella typhimurium secreted invasion determinants are homologous to Shigella Ipa proteins
JOURNAL Mol. Microbiol. 18 (3), 479-490 (1995)
MEDLINE 96342378
REFERENCE 2 (bases 1 to 3441)
AUTHORS Miller,S.I.
TITLE Direct Submission
JOURNAL Submitted (28-JUN-1995) Samuel I. Miller, Infectious Disease Unit,

```
          Massachusetts General Hospital, Harvard Medical School, Fruit
          Street, Boston, MA 02114, USA
FEATURES             Location/Qualifiers
     source          1. .3441
                     /organism="Salmonella typhimurium"
                     /strain="SL1344"
                     /db_xref="taxon:602"
                     /map="59 min"
     gene            1. .870
                     /gene="sspB"
     CDS             <1. .870
                     /gene="sspB"
                     /note="secreted protein"
                     /codon_start=1
                     /transl_table=11
                     /function="invasion determinant"
                     /product="SspB"
                     /protein_id="AAC43546.1"
                     /db_xref="PID:g1117812"
                     /db_xref="GI:1117812"
                     /translation="RKAEETNRIMGCIGKVLGALLTIVSVVAAVFTGGASLALAAVGL
                     AVMVADEIVKAATGVSFIQQALNPIMEHVLKPLMELIGKAITKALEGLGVDKKTAEMA
                     GSIVGAIVAAIAMVAVIVVVAVVGKGAAAKLGNALSKMMGETIKKLVPNVLKQLAQNG
                     SKLFTQGMQRITSGLGNVGSKMGLQTNALSKELVGNTLNKVALGMEVTNTAAQSAGGV
                     AEGVFIKNASEALADFMLARFAMDQIQQWLKQSVEIFGENQKVTAELQKAMSSAVQQN
                     ADASRFILRQSRA"
     gene            898. .2127
                     /gene="sspC"
     CDS             898. .2127
                     /gene="sspC"
                     /note="secreted protein"
                     /codon_start=1
                     /transl_table=11
                     /function="invasion determinant"
                     /evidence=experimental
                     /product="SspC"
                     /protein_id="AAC43547.1"
                     /db_xref="PID:g1117813"
                     /db_xref="GI:1117813"
                     /translation="MLISNVGINPAAYLNNHSVENSSQTASQSVSAKDILNSIGISSS
                     KVSDLGLSPTLSAPAPGVLTQTPGTITSFLKASIQNTDMNQDLNALANNVTTKANEVV
                     QTQLREQQAEVGKFFDISGMSSSAVALLAAANTLMLTLNQADSKLSGKLSLVSFDAAK
                     TTASSMMREGMNALSGSISQSALQLGITGVGAKLEYKGLQNERGALKHNAAKIDKLTT
                     ESHSIKNVLNGQNSVKLGAEGVDSLKSLNMKKTGTDATKNLNDATLKSNAGTSATESL
                     GIKDSNKQISPEHQAILSKRLESVESDIRLEQNTMDMTRIDARKMQMTGDLIMKNSVT
                     VGGIAGASGQYAATQERSEQQISQVNNRVASTASDEARESSRKSTSLIQEMLKTMESI
                     NQSKASALAAIAGNIRA"
     gene            2198. .3229
                     /gene="sspD"
     CDS             2198. .3229
                     /gene="sspD"
                     /note="secreted protein"
                     /codon_start=1
                     /transl_table=11
                     /function="invasion determinant"
                     /evidence=experimental
                     /product="SspD"
                     /protein_id="AAC43548.1"
                     /db_xref="PID:g1117814"
                     /db_xref="GI:1117814"
                     /translation="MLNIQNYSASPHPGIVAERPQTPSASEHVETAVVPSTTEHRGTD
                     IISLSQAATKIHQAQQTLQSTPPISEENNDERTLARQQLTSSLNALAKSGVSLSAEQN
                     ENLRSAFSAPTSALFSASPMAQPRTTISDAEIWDMVSQNISAIGDSYLGVYENVVAVY
                     TDFYQAFSDILSKMGGWLLPGKDGNTVKLDVTSLKNDLNSLVNKYNQINSNTVLFPAQ
                     SGSGVKVATEAEARQWLSELNLPNSCLKSYGSGYVVTVDLTPLQKMVQDIDGLGAPGK
                     DSKLEMDNAKYQAWQSGFKAQEENMKTTLQTLTQKYSNANSLYDNLVKVLSSTISSSL
                     ETAKSFLQG"
     gene            3248. .3441
                     /gene="sspA"
     CDS             3248. .>3441
                     /gene="sspA"
                     /note="secreted protein"
                     /codon_start=1
                     /transl_table=11
                     /function="unknown"
                     /product="SspA"
                     /protein_id="AAC43549.1"
                     /db_xref="PID:g1117815"
                     /db_xref="GI:1117815"
                     /translation="MVTSVRTQPPVIMPGMQTEIKTQATNLAANLSAVRESATATLSG
                     EIKGPQLEDFPALIKQASLDA"
BASE COUNT      978 a    759 c    925 g    779 t
ORIGIN


  Query Match           100.0%;  Score 1230;  DB 2;  Length 3441;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1230;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ATGTTAATTAGTAATGTGGGAATAAATCCCGCCGCTTATTTAAATAATCATTCTGTTGAG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    898 ATGTTAATTAGTAATGTGGGAATAAATCCCGCCGCTTATTTAAATAATCATTCTGTTGAG 957

Qy     61 AATAGTTCACAGACAGCTTCGCAATCCGTTAGCGCTAAAGATATTCTGAATAGTATTGGT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    958 AATAGTTCACAGACAGCTTCGCAATCCGTTAGCGCTAAAGATATTCTGAATAGTATTGGT 1017

Qy    121 ATTAGCAGCAGTAAAGTCAGTGACCTGGGGTTGAGTCCTACACTGAGCGCGCCTGCGCCA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1018 ATTAGCAGCAGTAAAGTCAGTGACCTGGGGTTGAGTCCTACACTGAGCGCGCCTGCGCCA 1077

Qy    181 GGGGTATTAACGCAAACCCCCGGAACGATCACGTCCTTTTTAAAAGCCAGTATTCAAAAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1078 GGGGTATTAACGCAAACCCCCGGAACGATCACGTCCTTTTTAAAAGCCAGTATTCAAAAT 1137

Qy    241 ACCGACATGAATCAGGATTTGAATGCTCTGGCAAATAATGTCACGACTAAAGCGAATGAG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1138 ACCGACATGAATCAGGATTTGAATGCTCTGGCAAATAATGTCACGACTAAAGCGAATGAG 1197

Qy    301 GTTGTGCAAACCCAGTTACGCGAGCAGCAGGCAGAAGTCGGAAAGTTTTTTGATATTAGC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1198 GTTGTGCAAACCCAGTTACGCGAGCAGCAGGCAGAAGTCGGAAAGTTTTTTGATATTAGC 1257

Qy    361 GGAATGTCTTCCAGTGCCGTTGCGCTGTTGGCTGCCGCGAATACGTTAATGCTGACGTTG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1258 GGAATGTCTTCCAGTGCCGTTGCGCTGTTGGCTGCCGCGAATACGTTAATGCTGACGTTG 1317

Qy    421 AACCAGGCTGATAGCAAACTGTCTGGTAAGTTGTCATTAGTCAGTTTTGATGCAGCTAAA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1318 AACCAGGCTGATAGCAAACTGTCTGGTAAGTTGTCATTAGTCAGTTTTGATGCAGCTAAA 1377

Qy    481 ACGACGGCAAGCTCCATGATGCGCGAAGGGATGAATGCGTTGTCCGGTAGTATTTCCCAG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1378 ACGACGGCAAGCTCCATGATGCGCGAAGGGATGAATGCGTTGTCCGGTAGTATTTCCCAG 1437

Qy    541 AGCGCGCTTCAGTTGGGGATCACTGGCGTGGGCGCCAAACTGGAATATAAGGGGCTGCAG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1438 AGCGCGCTTCAGTTGGGGATCACTGGCGTGGGCGCCAAACTGGAATATAAGGGGCTGCAG 1497

Qy    601 AATGAAAGAGGCGCGCTTAAACATAATGCCGCGAAGATCGATAAACTGACCACTGAAAGC 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1498 AATGAAAGAGGCGCGCTTAAACATAATGCCGCGAAGATCGATAAACTGACCACTGAAAGC 1557

Qy    661 CACAGTATTAAAAACGTGCTGAACGGGCAGAATAGCGTCAAACTCGGTGCTGAAGGCGTC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1558 CACAGTATTAAAAACGTGCTGAACGGGCAGAATAGCGTCAAACTCGGTGCTGAAGGCGTC 1617

Qy    721 GATTCTCTGAAATCGTTAAATATGAAGAAAACCGGTACCGATGCGACGAAAAATCTTAAT 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1618 GATTCTCTGAAATCGTTAAATATGAAGAAAACCGGTACCGATGCGACGAAAAATCTTAAT 1677

Qy    781 GATGCGACGCTTAAATCTAATGCCGGAACCAGCGCCACGGAAAGTCTGGGTATTAAAGAC 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1678 GATGCGACGCTTAAATCTAATGCCGGAACCAGCGCCACGGAAAGTCTGGGTATTAAAGAC 1737

Qy    841 AGTAATAAACAAATCTCCCCTGAACATCAGGCTATTCTGTCGAAACGTCTTGAGTCTGTC 900
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1738 AGTAATAAACAAATCTCCCCTGAACATCAGGCTATTCTGTCGAAACGTCTTGAGTCTGTC 1797

Qy     901 GAATCCGATATTCGTCTTGAGCAGAATACCATGGATATGACCCGAATCGATGCGCGCAAG 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1798 GAATCCGATATTCGTCTTGAGCAGAATACCATGGATATGACCCGAATCGATGCGCGCAAG 1857

Qy     961 ATGCAGATGACGGGCGATCTGATTATGAAGAACTCGGTCACGGTCGGTGGTATTGCAGGG 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1858 ATGCAGATGACGGGCGATCTGATTATGAAGAACTCGGTCACGGTCGGTGGTATTGCAGGG 1917

Qy    1021 GCGTCCGGGCAGTACGCCGCTACTCAGGAACGTTCCGAGCAGCAAATTAGCCAGGTGAAT 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1918 GCGTCCGGGCAGTACGCCGCTACTCAGGAACGTTCCGAGCAGCAAATTAGCCAGGTGAAT 1977

Qy    1081 AACCGGGTTGCCAGCACCGCATCGGACGAAGCCCGTGAAAGTTCACGTAAATCGACCAGC 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1978 AACCGGGTTGCCAGCACCGCATCGGACGAAGCCCGTGAAAGTTCACGTAAATCGACCAGC 2037

Qy    1141 CTGATTCAGGAAATGCTGAAAACAATGGAGAGCATTAACCAGTCGAAAGCATCCGCACTC 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2038 CTGATTCAGGAAATGCTGAAAACAATGGAGAGCATTAACCAGTCGAAAGCATCCGCACTC 2097

Qy    1201 GCTGCTATCGCAGGCAATATTCGCGCTTAA 1230
          ||||||||||||||||||||||||||||||
Db    2098 GCTGCTATCGCAGGCAATATTCGCGCTTAA 2127


RESULT   2
STSIPGNS
LOCUS       STSIPGNS     5393 bp    DNA             BCT       04-OCT-1995
DEFINITION  S.typhi spaT, sipB, sipC, sipD and sipA genes.
ACCESSION   X82670
NID         g1009001
VERSION     X82670.1  GI:1009001
KEYWORDS    sipA gene; sipB gene; sipC gene; sipD gene; spaT gene.
SOURCE      Salmonella typhi.
  ORGANISM  Salmonella typhi
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Salmonella.
REFERENCE   1  (bases 1 to 5393)
  AUTHORS   Hermant,D., Menard,R., Arricau,N., Parsot,C. and Popoff,M.Y.
  TITLE     Functional conservation of the Salmonella and Shigella effectors of
            entry into epithelial cells
  JOURNAL   Mol. Microbiol. 17 (4), 781-789 (1995)
  MEDLINE   96111497
REFERENCE   2  (bases 1 to 5393)
  AUTHORS   Popoff,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-1994) M. Popoff, Institut Pasteur, Unite des
            Enterobacteries, 28 rue du Docteur Roux, 75724 Paris Cedex 15,
            FRANCE
FEATURES             Location/Qualifiers
     source          1. .5393
                     /organism="Salmonella typhi"
                     /strain="Ty2"
                     /db_xref="taxon:601"
     gene            43. .540
                     /gene="spaT"
     CDS             43. .540
                     /gene="spaT"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA57987.1"
                     /db_xref="PID:g1009002"
                     /db_xref="GI:1009002"
                     /db_xref="SPTREMBL:Q57260"
                     /translation="MDYQNNVSEERVAEMIWDAVSEGATLKDVHGIPQDMMDGLYAHA
                     YEFYNQGRLDEAETFFRFLCIYDFYNPDYTMGLAAVCQLKKQFQKACDLYAVAFTLLK
                     NDYRPVFFTGQCQLLMRKAAKARQCFELVNERTEDESLRAKALVYLEALKTAETEQHS
                     EQEKE"
     gene            543. .2324
                     /gene="sipB"
     CDS             543. .2324
                     /gene="sipB"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA57988.1"
                     /db_xref="PID:g1009003"
                     /db_xref="GI:1009003"
                     /db_xref="SPTREMBL:Q56134"
                     /translation="MVNDASSISRSGYTQNPRLAEAAFEGVRKNTDFLKAADKAFKDV
                     VATKAGDLKAGTKSGESAINTVGLKPPTDAAREKLSSEGQLTLLLGKLMILLGDVSLS
                     QLESRLAVWQAMIESQKEMGIQVSKEFQTALGEAQEATDLYEASIKKTDTAKSVYDAA
                     AKKLTQAQNKLQSLDPADPGYAQAEAAVEQAGKEATEAKEALDKATDAIVKAGTDAKA
                     KAEKADNILTKFQGTANAASQNQVSQGEQDNLSNVARLTMLMAMFIEIVGKNTEESLQ
                     NDLALFNALQEGRQAEMEKKSAEFQEETRKAEETNRIMGCIGKVLGALLTIVSVVAAV
                     FTGGASLALAAVGLAVMVADEIVKAATGVSFIQQALNPIMEHVLKPLMELIGKAITKA
                     LEGLGVDKKTAEMAGSIVGAIVAAIAMVAVIVVVAVVGKGAAAKLGNALSKMMGETIK
                     KLVPNVLKQLAQNGSKLFTQGMQRITSGLGNVGSKMGLQTNALSKELVGNTLNKVALG
                     MEVTNTAAQSAGGVAEGVFIKNASEALADFMLARFAMDQIQQWLKQSVEIFGENQKVT
                     AELQKAMSSAVQQNADASRFILRQSRA"
     gene            2352. .3581
                     /gene="sipC"
     CDS             2352. .3581
                     /gene="sipC"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA57989.1"
                     /db_xref="PID:g1009004"
                     /db_xref="GI:1009004"
                     /db_xref="SPTREMBL:Q56135"
                     /translation="MLISNVGINPAAYLNNHSVENSSQTASQSVSAKDILNSIGISSS
                     KVSDLGLSPTLSAPAPGVLTQTPGTITSFLKASIQNTDMNQDLNALANNVTTKANEVV
                     QTQLREQQAEVGKFFDISGMSSSAVALLAAANTLMLTLNQADSKLSGKLSLVSFDAAK
                     TTASSMMREGMNALSGSISQSALQLGITGVGAKLEYKGLQNERGALKHNAAKIDKLTT
                     ESHSIKNVLNGQNSVKLGAEGVDSLKSLNMKKTGTDATKNLNDATLKSNAGTSATESL
                     GIKNSNKQISPEHQAILSKRLESVESDIRLEQNTMDMTRIDARKMQMTGDLIMKNSVT
                     VGGIAGASRQYAATQERSEQQISQVNNRVASTASDEARESSRKSTSLIQEMLKTMESI
                     NQSKASALAAIAGNIRA"
     gene            3651. .4670
                     /gene="sipD"
     CDS             3651. .4670
                     /gene="sipD"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA57990.1"
                     /db_xref="PID:g1009005"
                     /db_xref="GI:1009005"
                     /db_xref="SPTREMBL:Q56136"
                     /translation="MLNIQNYSASPHPGIVAERPQTPSASEHAEIAVVPSTTEHRGTD
                     IISLSQAATKIQQAQQTLQSTPPISEENNDERTLARQQLTSSLNALAKSGVSLSAEQN
                     ENLRSTFSARRRPYLALRLWPARTTISDAEIWDMVSQNISAIGDSYLGVYENVVAVYT
                     DFYQAFSDILSKMGGWLSPGKDGNTIKLNVDSLKSEISSLINKYTQINKNTILFPSQT
                     GSGMTTATKAEAEQWIKELNLPDSCLKASGSGYVVLVDTGPLSKMVSDLNGIGSGSAL
                     ELDNAKYQAWQSGFKAQEENLKTTLQTLTQKYSNANSLYDNLVKVLSSTISSSLETAK
                     SFLQG"
     gene            4689. .5393
                     /gene="sipA"
     CDS             4689. .>5393
                     /gene="sipA"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA57991.1"
                     /db_xref="PID:g1009006"
                     /db_xref="GI:1009006"
                     /db_xref="SPTREMBL:Q56137"
                     /translation="MVTSVRTQPPVIMPGMQTEIKTQAINLAANLSAVRESATATLSG
                     EIKGQQLEDFPALIKQASLDALFKCGKDAEALKEVFTNSNNVAGKKAIMEFAGLFRSA
                     LNATSDSPEAKTLLMKVGAEYTAQIIKDGLKEKSAFGPWLPETKKAEAKLENLEKQLL
                     DIIKNNTGGELSKLSTNLVMQEVMPYIASCIEHNFGCTLDPLTRSSLTQLVDKAAAKA
                     VEALDMCHQKLTQEQGT"
BASE COUNT     1584 a    1156 c    1458 g    1195 t
ORIGIN
```

```
  Query Match             99.3%;  Score 1222;  DB 1;  Length 5393;
  Best Local Similarity   99.6%;  Pred. No. 0;
  Matches 1225;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 ATGTTAATTAGTAATGTGGGAATAAATCCCGCCGCTTATTTAAATAATCATTCTGTTGAG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2352 ATGTTAATTAGTAATGTGGGAATAAATCCCGCCGCTTATTTAAATAATCATTCTGTTGAG 2411

Qy      61 AATAGTTCACAGACAGCTTCGCAATCCGTTAGCGCTAAAGATATTCTGAATAGTATTGGT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2412 AATAGTTCACAGACAGCTTCGCAATCCGTTAGCGCTAAAGATATTCTGAATAGTATTGGT 2471

Qy     121 ATTAGCAGCAGTAAAGTCAGTGACCTGGGGTTGAGTCCTACACTGAGCGCGCCTGCGCCA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2472 ATTAGCAGCAGTAAAGTCAGTGACCTGGGGTTGAGTCCTACACTGAGCGCGCCTGCGCCA 2531

Qy     181 GGGGTATTAACGCAAACCCCCGGAACGATCACGTCCTTTTTAAAAGCCAGTATTCAAAAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2532 GGGGTATTAACGCAAACCCCCGGAACGATCACGTCCTTTTTAAAAGCCAGTATTCAAAAT 2591

Qy     241 ACCGACATGAATCAGGATTTGAATGCTCTGGCAAATAATGTCACGACTAAAGCGAATGAG 300
          ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db    2592 ACCGACATGAATCAGGATTTGAATGCCCTGGCAAATAATGTCACGACTAAAGCGAATGAG 2651

Qy     301 GTTGTGCAAACCCAGTTACGCGAGCAGCAGGCAGAAGTCGGAAAGTTTTTTGATATTAGC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2652 GTTGTGCAAACCCAGTTACGCGAGCAGCAGGCAGAAGTCGGAAAGTTTTTTGATATTAGC 2711

Qy     361 GGAATGTCTTCCAGTGCCGTTGCGCTGTTGGCTGCCGCGAATACGTTAATGCTGACGTTG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2712 GGAATGTCTTCCAGTGCCGTTGCGCTGTTGGCTGCCGCGAATACGTTAATGCTGACGTTG 2771

Qy     421 AACCAGGCTGATAGCAAACTGTCTGGTAAGTTGTCATTAGTCAGTTTTGATGCAGCTAAA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2772 AACCAGGCTGATAGCAAACTGTCTGGTAAGTTGTCATTAGTCAGTTTTGATGCAGCTAAA 2831

Qy     481 ACGACGGCAAGCTCCATGATGCGCGAAGGGATGAATGCGTTGTCCGGTAGTATTTCCCAG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2832 ACGACGGCAAGCTCCATGATGCGCGAAGGGATGAATGCGTTGTCCGGTAGTATTTCCCAG 2891

Qy     541 AGCGCGCTTCAGTTGGGGATCACTGGCGTGGGCGCCAAACTGGAATATAAGGGGCTGCAG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2892 AGCGCGCTTCAGTTGGGGATCACTGGCGTGGGCGCCAAACTGGAATATAAGGGGCTGCAG 2951

Qy     601 AATGAAAGAGGCGCGCTTAAACATAATGCCGCGAAGATCGATAAACTGACCACTGAAAGC 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2952 AATGAAAGAGGCGCGCTTAAACATAATGCCGCGAAGATCGATAAACTGACCACTGAAAGC 3011

Qy     661 CACAGTATTAAAAACGTGCTGAACGGGCAGAATAGCGTCAAACTCGGTGCTGAAGGCGTC 720
          |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    3012 CACAGTATTAAAAACGTGCTGAACGGGCAGAATAGCGTCAAACTTGGTGCTGAAGGCGTC 3071

Qy     721 GATTCTCTGAAATCGTTAAATATGAAGAAAACCGGTACCGATGCGACGAAAAATCTTAAT 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3072 GATTCTCTGAAATCGTTAAATATGAAGAAAACCGGTACCGATGCGACGAAAAATCTTAAT 3131

Qy     781 GATGCGACGCTTAAATCTAATGCCGGAACCAGCGCCACGGAAAGTCTGGGTATTAAAGAC 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  ||
Db    3132 GATGCGACGCTTAAATCTAATGCCGGAACCAGCGCCACGGAAAGTCTGGGTATTAAAAAC 3191

Qy     841 AGTAATAAACAAATCTCCCCTGAACATCAGGCTATTCTGTCGAAACGTCTTGAGTCTGTC 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3192 AGTAATAAACAAATCTCCCCTGAACATCAGGCTATTCTGTCGAAACGTCTTGAGTCTGTC 3251

Qy     901 GAATCCGATATTCGTCTTGAGCAGAATACCATGGATATGACCCGAATCGATGCGCGCAAG 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3252 GAATCCGATATTCGTCTTGAGCAGAATACCATGGATATGACCCGAATCGATGCGCGCAAG 3311

Qy     961 ATGCAGATGACGGGCGATCTGATTATGAAGAACTCGGTCACGGTCGGTGGTATTGCAGGG 1020
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    3312 ATGCAGATGACGGGCGATCTGATTATGAAGAACTCAGTCACGGTCGGTGGTATTGCAGGG 3371

Qy    1021 GCGTCCGGGCAGTACGCCGCTACTCAGGAACGTTCCGAGCAGCAAATTAGCCAGGTGAAT 1080
          |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3372 GCGTCCAGGCAGTACGCCGCTACTCAGGAACGTTCCGAGCAGCAAATTAGCCAGGTGAAT 3431

Qy    1081 AACCGGGTTGCCAGCACCGCATCGGACGAAGCCCGTGAAAGTTCACGTAAATCGACCAGC 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3432 AACCGGGTTGCCAGCACCGCATCGGACGAAGCCCGTGAAAGTTCACGTAAATCGACCAGC 3491

Qy    1141 CTGATTCAGGAAATGCTGAAAACAATGGAGAGCATTAACCAGTCGAAAGCATCCGCACTC 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3492 CTGATTCAGGAAATGCTGAAAACAATGGAGAGCATTAACCAGTCGAAAGCATCCGCACTC 3551

Qy    1201 GCTGCTATCGCAGGCAATATTCGCGCTTAA 1230
          ||||||||||||||||||||||||||||||
Db    3552 GCTGCTATCGCAGGCAATATTCGCGCTTAA 3581
```

RESULT 3
STU25631

```
LOCUS       STU25631     3622 bp    DNA             BCT       06-SEP-1995
DEFINITION  Salmonella typhimurium invasin homologs SicA (sicA), SipB (sipB)
            and SipC  (sipC) genes, complete cds.
ACCESSION   U25631
NID         g975293
VERSION     U25631.1  GI:975293
KEYWORDS    .
SOURCE      Salmonella typhimurium.
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Salmonella.
REFERENCE   1  (bases 1 to 3622)
  AUTHORS   Kaniga,K., Tucker,S., Trollinger,D. and Galan,J.E.
  TITLE     Homologs of the Shigella IpaB and IpaC invasins are required for
            Salmonella typhimurium entry into cultured epithelial cells
  JOURNAL   J. Bacteriol. 177 (14), 3965-3971 (1995)
  MEDLINE   95332200
REFERENCE   2  (bases 1 to 3622)
  AUTHORS   Galan,J.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-APR-1995) Jorge E. Galan, Molecular Genetics and
            Microbiology, School of Medicine, SUNY Stony Brook, Stony Brook, NY
            11794-5222, USA
FEATURES             Location/Qualifiers
     source          1. .3622
                     /organism="Salmonella typhimurium"
                     /strain="SL1344"
                     /db_xref="taxon:602"
     gene            75. .572
                     /gene="sicA"
     CDS             75. .572
                     /gene="sicA"
                     /codon_start=1
                     /transl_table=11
                     /product="SicA"
                     /protein_id="AAA75168.1"
                     /db_xref="PID:g975294"
                     /db_xref="GI:975294"
                     /translation="MDYQNNVSEERVAEMIWDAVSEGATLKDVHGIPQDMMDGLYAHA
                     YEFYNQGRLDEAETFFRFLCIYDFYNPDYTMGLAAVCQLKKQFQKACDLYAVAFTLLK
                     NDYRPVFFTGQCQLLMRKAAKARQCFELVNERTEDESLRAKALVYLEALKTAETEQHS
                     EQEKE"
     gene            575. .2356
                     /gene="sipB"
     CDS             575. .2356
                     /gene="sipB"
                     /codon_start=1
                     /transl_table=11
                     /product="SipB"
                     /protein_id="AAA75169.1"
                     /db_xref="PID:g975295"
                     /db_xref="GI:975295"
                     /translation="MVNDASSISRSGYTQNPRLAEAAFEGVRKNTDFLKAADKAFKDV
                     VATKAGDLKAGTKSGESAINTVGLKPPTDAAREKLSSEGQLTLLLGKLMTLLGDVSLS
                     QLESRLAVWQAMIESQKEMGIQVSKEFQTALGEAQEATDLYEASIKKTDTAKSVYDAA
```

```
                     TKKLTQAQNKLQSLDPADPGYAQAEAAVEQAGKEATEAKEALDKATDATVKAGTDAKA
                     KAEKADNILTKFQGTANAASQNQVSQGEQDNLSNVARLTMLMAMFIEIVGKNTEESLQ
                     NDLALFNALQEGRQAEMEKKSAEFQEETRKAEETNRIMGCIGKVLGALLTIVSVVAAV
                     FTGGASLALAAVGLAVMVADEIVKAATGVSFIQQALNPIMEHVLKPLMELIGKAITKA
                     LEGLGVDKKTAEMAGSIVGAIVAAIAMVAVIVVVAVVGKGAAAKLGNALSKMMGETIK
                     KLVPNVLKQLAQNGSKLFTQGMQRITSGLGNVGSKMGLQTNALSKELVGNTLNKVALG
                     MEVTNTAAQSAGGVAEGVFIKNASEALADFMLARFAMDQIQQWLKQSVEIFGENQKVT
                     AELQKAMSSAVQQNADASRFILRQSRA"
     gene            2384. .3613
                     /gene="sipC"
     CDS             2384. .3613
                     /gene="sipC"
                     /codon_start=1
                     /transl_table=11
                     /product="SipC"
                     /protein_id="AAA75170.1"
                     /db_xref="PID:g975296"
                     /db_xref="GI:975296"
                     /translation="MLISNVGINPAAYLNNHSVENSSQTASQSVSAKDILNSIGISSS
                     KVSDLGLSPTLSAPAPGVLTQTPGTITSSLKASIQNTDMNQDLNALANNVTTKANEVV
                     QTQLREQQAEVGKFFDISGMSSSAVALLAAANTLMLTLNQADSKLSGKLSLVSFDAAK
                     TTASSMMREGMNALSGSISQSALQLGITGVGAKLEYKGLQNERGALKHNAAKIDKLTT
                     ESHSIKNVLNGQNSVKLGAEGVDSLKSLNIRKPVPMRRKILMMRRLNLMPEPAPRKVW
                     VLKTVINKVSLNIYILSKRLESVESDIRLEQNYMDITRIDSAQDADDGRSDYEELGHG
                     RWYCRGVRAVRRYSGNVSEQQISQVNNRVASTASDEARESSRKSTSLIQEMLKTMESI
                     NQSKASALAAIAGNIRA"
BASE COUNT     1049 a    759 c   1008 g    806 t
ORIGIN
```

Query Match 94.1%; Score 1157; DB 1; Length 3622;
Best Local Similarity 98.3%; Pred. No. 0;
Matches 1212; Conservative 0; Mismatches 15; Indels 6; Gaps 4;

```
Qy       1 ATGTTAATTAGTAATGTGGGAATAAATCCCGCCGCTTATTTAAATAATCATTCTGTTGAG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2384 ATGTTAATTAGTAATGTGGGAATAAATCCCGCCGCTTATTTAAATAATCATTCTGTTGAG 2443

Qy      61 AATAGTTCACAGACAGCTTCGCAATCCGTTAGCGCTAAAGATATTCTGAATAGTATTGGT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2444 AATAGTTCACAGACAGCTTCGCAATCCGTTAGCGCTAAAGATATTCTGAATAGTATTGGT 2503

Qy     121 ATTAGCAGCAGTAAAGTCAGTGACCTGGGGTTGAGTCCTACACTGAGCGCGCCTGCGCCA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2504 ATTAGCAGCAGTAAAGTCAGTGACCTGGGGTTGAGTCCTACACTGAGCGCGCCTGCGCCA 2563

Qy     181 GGGGTATTAACGCAAACCCCCCGGAACGATCACGTCCTTTTTAAAAGCCAGTATTCAAAAT 240
          |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    2564 GGGGTATTAACGCAAACCCCCGGAACGATCACGTCCTCTTTAAAAGCCAGTATTCAAAAT 2623

Qy     241 ACCGACATGAATCAGGATTTGAATGCTCTGGCAAATAATGTCACGACTAAAGCGAATGAG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2624 ACCGACATGAATCAGGATTTGAATGCTCTGGCAAATAATGTCACGACTAAAGCGAATGAG 2683

Qy     301 GTTGTGCAAACCCAGTTACGCGAGCAGCAGGCAGAAGTCGGAAAGTTTTTTGATATTAGC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2684 GTTGTGCAAACCCAGTTACGCGAGCAGCAGGCAGAAGTCGGAAAGTTTTTTGATATTAGC 2743

Qy     361 GGAATGTCTTCCAGTGCCGTTGCGCTGTTGGCTGCCGCGAATACGTTAATGCTGACGTTG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2744 GGAATGTCTTCCAGTGCCGTTGCGCTGTTGGCTGCCGCGAATACGTTAATGCTGACGTTG 2803

Qy     421 AACCAGGCTGATAGCAAACTGTCTGGTAAGTTGTCATTAGTCAGTTTTGATGCAGCTAAA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2804 AACCAGGCTGATAGCAAACTGTCTGGTAAGTTGTCATTAGTCAGTTTTGATGCAGCTAAA 2863

Qy     481 ACGACGGCAAGCTCCATGATGCGCGAAGGGATGAATGCGTTGTCCGGTAGTATTTCCCAG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2864 ACGACGGCAAGCTCCATGATGCGCGAAGGGATGAATGCGTTGTCCGGTAGTATTTCCCAG 2923

Qy     541 AGCGCGCTTCAGTTGGGGATCACTGGCGTGGGCGCCAAACTGGAATATAAGGGGCTGCAG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2924 AGCGCGCTTCAGTTGGGGATCACTGGCGTGGGCGCCAAACTGGAATATAAGGGGCTGCAG 2983

Qy     601 AATGAAAGAGGCGCGCTTAAACATAATGCCGCGAAGATCGATAAACTGACCACTGAAAGC 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2984 AATGAAAGAGGCGCGCTTAAACATAATGCCGCGAAGATCGATAAACTGACCACTGAAAGC 3043

Qy     661 CACAGTATTAAAAACGTGCTGAACGGGCAGAATAGCGTCAAACTCGGTGCTGAAGGCGTC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3044 CACAGTATTAAAAACGTGCTGAACGGGCAGAATAGCGTCAAACTCGGTGCTGAAGGCGTC 3103

Qy     721 GATTCTCTGAAATCGTTAAATATGAAGAAAACCGGTACCGATGCGACGAAAAATCTTAAT 780
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    3104 GATTCTCTGAAATCGTTAAATAT-AAGAAAACCGGTACCGATGCGACGAAAAATCTTAAT 3162

Qy     781 GATGCGACGCTTAAATCTAATGCCGGAACCAGCGCCACGGAAAGTCTGGGTATTAAAGAC 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3163 GATGCGACGCTTAAATCTAATGCCGGAACCAGCGCCACGGAAAGTCTGGGTATTAAAGAC 3222

Qy     841 AGTAATAAACAAATCTCCCCTGAACATCAGGCTATTCTGTCGAAACGTCTTGAGTCTGTC 900
          |||||||||||||    ||||||||||    ||| ||||||||||||||||||||||||||
Db    3223 AGTAATAAACAAAGTCTCCCTGAACATC--TATATCTTGTCGAAACGTCTTGAGTCTGTC 3280

Qy     901 GAATCCGATATTCGTCTTGAGCAGAATACCATGGATATGACCCGAATCGAT-GCGCGCAA 959
          ||||||||||||||||||||||||||  ||||||||| |||||||||||| ||||||||
Db    3281 GAATCCGATATTCGTCTTGAGCAGAATTACATGGATATTACCCGAATCGATAGCGCGCAA 3340

Qy     960 GATGCAGATGACGGGCGATCTGATTATGAAGAACTCGGTCACGGTCGGTGGTATTGCAGG 1019
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3341 GATGCAGATGACGGGCGATCTGATTATGAAGAACTCGGTCACGGTCGGTGGTATTGCAGG 3400

Qy    1020 GGCGTCCGGGCAGTACGCCGCTACTCAGG--AACGTTCCGAGCAGCAAATTAGCCAGGTG 1077
          |||||||||||||||||||||||||||||  |  ||||||||||||||||||||||||||
Db    3401 GGCGTCCGGGCAGTACGCCGCTACTCAGGTAACGTTTCCGAGCAGCAAATTAGCCAGGTG 3460

Qy    1078 AATAACCGGGTTGCCAGCACCGCATCGGACGAAGCCCGTGAAAGTTCACGTAAATCGACC 1137
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3461 AATAACCGGGTTGCCAGCACCGCATCGGACGAAGCCCGTGAAAGTTCACGTAAATCGACC 3520

Qy    1138 AGCCTGATTCAGGAAATGCTGAAAACAATGGAGAGCATTAACCAGTCGAAAGCATCCGCA 1197
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3521 AGCCTGATTCAGGAAATGCTGAAAACAATGGAGAGCATTAACCAGTCGAAAGCATCCGCA 3580

Qy    1198 CTCGCTGCTATCGCAGGCAATATTCGCGCTTAA 1230
          ||||||||||||||||||||||||||||||||
Db    3581 CTCGCTGCTATCGCAGGCAATATTCGCGCTTAA 3613
```

```
RESULT   4
SDU66877
LOCUS       SDU66877     2995 bp    DNA             BCT       29-AUG-1996
DEFINITION  Salmonella dublin SicA (sicA) and SipB (sipB) genes, complete cds.
ACCESSION   U66877
NID         g1513311
VERSION     U66877.1  GI:1513311
KEYWORDS    .
SOURCE      Salmonella dublin.
  ORGANISM  Salmonella dublin
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Salmonella.
REFERENCE   1  (bases 1 to 2995)
  AUTHORS   Mullan,P.B., Gautier,A.V., Wood,M.W., Edwards,M.H., Jones,B.V. and
            Galyov,E.E.
  TITLE     Sip invasins of Salmonella dublin are involved in macrophage
            cytotoxicity and mouse virulence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2995)
  AUTHORS   Mullan,P.B., Gautier,A.V., Wood,M.W., Edwards,M.H., Jones,B.V. and
            Galyov,E.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-AUG-1996) Institute for Animal Health, Compton,
            Newbury, Berkshire, RG20 7NN, UK
FEATURES             Location/Qualifiers
     source          1. .2995
                     /organism="Salmonella dublin"
```

```
                     /strain="2229"
                     /db_xref="taxon:593"
     gene            131. .628
                     /gene="sicA"
     CDS             131. .628
                     /gene="sicA"
                     /codon_start=1
                     /transl_table=11
                     /product="SicA"
                     /protein_id="AAB06794.1"
                     /db_xref="PID:g1513312"
                     /db_xref="GI:1513312"
                     /translation="MDYQNNVSEERVAEMIWDAVSEGATLKDVHGIPQDMMDGLYAHA
                     YEFYNQGRLDEAETFFRFLCIYDFYNPDYTMGLAAVCQLKKQFQKACDLYAVAFTLLK
                     NDYRPVFFTGQCQLLMRKAAKARQCFELVNERTEDESLRAKALVYLEALKTAETEQHS
                     EQEKE"
     gene            631. .2412
                     /gene="sipB"
     CDS             631. .2412
                     /gene="sipB"
                     /function="virulence-associated secreted protein"
                     /codon_start=1
                     /transl_table=11
                     /product="SipB"
                     /protein_id="AAB06795.1"
                     /db_xref="PID:g1513313"
                     /db_xref="GI:1513313"
                     /translation="MVNDASSISRSGYTQNPRLAEAAFEGVRKNTDFLKAADKAFKDV
                     VATKAGDLKAGTKSGESAINTVGLKPPTDAAREKLSSEGQLTLLLGKLMTLLGDVSLS
                     QLESRLAVWQAMIESQKEMGIQVSKEFQTALGEAQEATDLYEASIKKTDTAKSVYDAA
                     TKKLTQAQNKLQSLDPADPGYAQAEAAVEQAGKEATEAKEALDKATDATVKAGTDAKA
                     KAEKADNILTKFQGTANAASQNQVSQGEQDNLSNVARLTMLMAMFIEIVGKNTEESLQ
                     NDLALFNALQEGRQAEMEKKSAEFQEETRKAEETNRIMGCIGKVLGALLTIVSVVAAV
                     FTGGASLALAAVGLAVMVADEIVKAATGVSFIQQALNPIMEHVLKPLMELIGKAITKA
                     LEGLGVDKKTAEMAGSIVGAIVAAIAMVAVIVVVAVVGKGAAAKLGNALSKMMGETIK
                     KLVPNVLKQLAQNGSKLFTQGMQRITSGLGNVGSKMGLQTNALSKELVGNTLNKVALG
                     MEVTNTAAQSAGGVAEGVFIKNASEALADFMLARFAMDQIQQWLKQSVEIFGENQKVT
                     AELQKAMSSAVQQNADASRFILRQSRA"
BASE COUNT      851 a    614 c    847 g    683 t
ORIGIN


  Query Match             44.8%;  Score 551.2;  DB 1;  Length 2995;
  Best Local Similarity   99.5%;  Pred. No. 4.1e-149;
  Matches  553;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 ATGTTAATTAGTAATGTGGGAATAAATCCCGCCGCTTATTTAAATAATCATTCTGTTGAG 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2440 ATGTTAATTAGTAATGTGGGAATAAATCCCGCCGCTTATTTAAATAATCATTCAGTTGAG 2499

Qy      61 AATAGTTCACAGACAGCTTCGCAATCCGTTAGCGCTAAAGATATTCTGAATAGTATTGGT 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2500 AATAGTTCACAGACAGCTTCGCAATCCGTTAGCGCTAAAGATATTCTGAATAGTATTGGT 2559

Qy     121 ATTAGCAGCAGTAAAGTCAGTGACCTGGGGTTGAGTCCTACACTGAGCGCGCCTGCGCCA 180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2560 ATTAGCAGCAGTAAAGTCAGTGACCTGGGGTTGAGTCCTACACTGAGCGCGCCTGCGCCA 2619

Qy     181 GGGGTATTAACGCAAACCCCCGGAACGATCACGTCCTTTTTAAAAGCCAGTATTCAAAAT 240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2620 GGGGTATTAACGCAAACCCCCGGAACGATCACGTCCTTTTTAAAAGCCAGTATTCAAAAT 2679

Qy     241 ACCGACATGAATCAGGATTTGAATGCTCTGGCAAATAATGTCACGACTAAAGCGAATGAG 300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2680 ACCGACATGAATCAGGATTTGAATGCTCTGGCAAATAATGTCACGACTAAAGCGAATGAG 2739

Qy     301 GTTGTGCAAACCCAGTTACGCGAGCAGCAGGCAGAAGTCGGAAAGTTTTTTGATATTAGC 360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2740 GTTGTGCAAACCCAGTTACGCGAGCAGCAGGCAGAAGTCGGAAAGTTTTTTGATATTAGC 2799

Qy     361 GGAATGTCTTCCAGTGCCGTTGCGCTGTTGGCTGCCGCGAATACGTTAATGCTGACGTTG 420
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2800 GGAATGTCTTCCAGTGCCGTTGCGCTGTTGGCTGCCGCGAATACGTTAATGCTGACGTTG 2859

Qy     421 AACCAGGCTGATAGCAAACTGTCTGGTAAGTTGTCATTAGTCAGTTTTGATGCAGCTAAA 480
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2860 AACCAGGCTGATAGCAAACTGTCTGGTAAGTTGTCATTAGTCAGCTTTGATGCAGCTAAA 2919

Qy     481 ACGACGGCAAGCTCCATGATGCGCGAAGGGATGAATGCGTTGTCCGGTAGTATTTCCCAG 540
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2920 ACGACAGCAAGCTCCATGATGCGCGAAGGGATGAATGCGTTGTCCGGTAGTATTTCCCAG 2979

Qy     541 AGCGCGCTTCAGTTGG 556
           ||||||||||||||||
Db    2980 AGCGCGCTTCAGTTGG 2995


RESULT   5
SHFIPABCD
LOCUS       SHFIPABCD    5255 bp    DNA             BCT       19-MAR-1991
DEFINITION  S.flexneri ipaBCD region encoding invasion plasmid antigens.
ACCESSION   J04117
NID         g152740
VERSION     J04117.1  GI:152740
KEYWORDS    ipaA gene; ipaB gene; ipaC gene; ipaD; membrane protein.
SOURCE      S.flexeri (strain M90T-W) DNA, clone pHC17.
  ORGANISM  Shigella flexneri
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Shigella.
REFERENCE   1  (bases 1 to 5255)
  AUTHORS   Venkatesan,M.M., Buysse,J.M. and Kopecko,D.J.
  TITLE     Characterization of invasion plasmid antigen genes (ipaBCD) from
            Shigella flexneri
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 85, 9317-9321 (1988)
  MEDLINE   89057927
COMMENT     Draft entry and clean copy of sequence [1] kindly submitted by
            M.M.Venkatesan, 03-JAN-1989.
FEATURES             Location/Qualifiers
     source          1. .5255
                     /organism="Shigella flexneri"
                     /db_xref="taxon:623"
     CDS             606. .1073
                     /note="18 kD protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA26521.1"
                     /db_xref="PID:g152741"
                     /db_xref="GI:152741"
                     /translation="MSLNITENESISTAVIDAINSGATLKDINAIPDDMMDDIYSYAY
                     DFYNKGRIEEAEVFFRFLCIYDFYNVDYIMGLAAIYQIKEQFQQAADLYAVAFALGKN
                     DYTPVFHTGQCQLRLKAPLKAKECFELVIQHSNDEKLKIKAQSYLDAIQDIKE"
     CDS             1079. .2821
                     /note="IpaB protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA26522.1"
                     /db_xref="PID:g152742"
                     /db_xref="GI:152742"
                     /translation="MHNVSTTTTGFPLAKILTSTELGDNTIQAANDAANKLFSLIIAD
                     LTANQNINTTNAHSTSNILIPELKAPKSLNASSQLTLLIGNLIQILGEKSLTALTNKI
                     TAWKSQQQARQQKNLEFSDKINTLLSETEGLTRDYEKQINKLKNADSKIKDLENKINQ
                     IQTRLSNLDPESPEKKKLSREEIQLTIKKDAAVKDRTLIEQKTLSIHSKLTDKSMQLE
                     KEIDSFSAFSNTASAEQLSTQQKSLTGLASVTQLMATFIQLVGKNNEESLKNDLALFQ
                     SLQESRKTEMERKSDEYAAEVRKAEELNRVMGCVGKILGALLTIVSVVAAAFSGGASL
                     ALAAVGLALMVTDAIVQAATGNSFMEQALNPIMKAVIEPLIKLLSDAFTKMLEGLGVD
                     SKKAKMIGSILGAIAGALVLVAAVVLVATVGKQAAAKLAENIGKIIGKTLTDLIPKFL
                     KNFSSQLDDLITNAVARLNKFLGAAGDEVISKQIISTHLNQAVLLGESVNSATQAGGS
                     VASAVFQNSASTNLADLTLSKYQVEQLSKYISEAIEKFGQLQEVIADLLASMSNSQAN
                     RTDVAKAILQQTTA"
     CDS             2784. .3932
                     /note="IpaC protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA26523.1"
                     /db_xref="PID:g152743"
                     /db_xref="GI:152743"
```

```
                     /translation="MLQKQFCNKLLLDTNKENVMEIQNTKPTQTLYTDISTKQTQSSS
                     ETQKSQNYQQIAAHIPLNVGKNPVLTTTLNDDQLLKLSEQVQHDSEIIARLTDKKMKD
                     LSEMSHTLTPENTLDISSLSSNAVSLIISVAVLLSALRTAETKLGSQLSLIAFDATKS
                     AAENIVRQGLAALSSSITGAVTQVGITGIGAKKTHSGISDQKGALRKNLATAQSLEKE
                     LAGSKLGLNKQIDTNITSPQTNSSTKFLGKNKLAPDNISLSTEHKTSLSSPDISLQDK
                     IDTQRRTYELNTLSAQQKQNIGRATMETSAVAGNISTSGGRYASALEEEEQLISQASS
                     KQAEEASQVSKEASQATNQLIQKLLNIIDSINQSKNSAASQIAGNIRA"
     CDS             3983. .4981
                     /note="IpaD protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA26524.1"
                     /db_xref="PID:g152744"
                     /db_xref="GI:152744"
                     /translation="MNITTLTNSISTSSFSPNNTNGSSTETVNSDIKTTTSSHPVSSL
                     TMLNDTLHNIRTTNQALKKELSQKTLTKTSLEEIALHSSQISMDVNKSAQLLDILSRN
                     EYPINKDARELLHSAPKEAELDGDQMISHRELWAKIANSINDINEQYLKVYEHAVSSY
                     TQMYQDFSAVLSSLAGWISPGGNDGNSVKLQVNSLKKALEELKEKYKDKPLYPANNTV
                     SQEQANKWLTELGGTIGKVSQKNGGYVVSINMTPIDNMLKSLDNLGGNGEVVLDNAKY
                     QAWNAGFSAEDETMKNNLQTLVQKYSNANSIFDNLVKVLSSTISSCTDTDKLFLHF"
     CDS             4990. .>5255
                     /note="IpaA protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA26525.1"
                     /db_xref="PID:g551981"
                     /db_xref="GI:551981"
                     /translation="MHNVNNTQAPTFLYKATSPSSTEYSELKSKISDIHSSQTSLKTP
                     ASVSEKENFATSFNQKCLDFLFSSSGKEDVLRSIYSNSMNAYAK"
BASE COUNT     1950 a   1037 c    889 g   1379 t
ORIGIN
```

```
  Query Match             5.4%;  Score 66.4;  DB 1;  Length 5255;
  Best Local Similarity   48.2%;  Pred. No. 1.3e-08;
  Matches  187;  Conservative    0;  Mismatches  201;  Indels    0;  Gaps    0;

Qy     843 TAATAAACAAATCTCCCCTGAACATCAGGCTATTCTGTCGAAACGTCTTGAGTCTGTCGA 902
           ||||| |    | ||  |||||||| |  ||  |||       |    |   ||| |  |
Db    3545 TAATATATCCCTGTCAACTGAACATAAAACTTCTCTTAGTTCTCCCGATATTTCTTTGCA 3604

Qy     903 ATCCGATATTCGTCTTGAGCAGAATACCATGGATATGACCCGAATCGATGCGCGCAAGAT 962
           | |||       ||   |  ||   || | |     |  |||||   | |
Db    3605 GGATAAAATTGACACCCAGAGAAGAACTTACGAGCTCAATACCCTTTCTGCGCAGCAAAA 3664

Qy     963 GCAGATGACGGGCGATCTGATTATGAAGAACTCGGTCACGGTCGGTGGTATTGCAGGGGC 1022
           || |  |  |||  |   |  ||| | |  || | |   |  |||  |||  |     |
Db    3665 ACAAAACATTGGCCGTGCAACAATGGAAACATCAGCCGTTGCTGGTAATATATCCACATC 3724

Qy    1023 GTCCGGGCAGTACGCCGCTACTCAGGAACGTTCCGAGCAGCAAATTAGCCAGGTGAATAA 1082
           ||||  || ||  || |||  |||       || || | ||| || ||||  |  |
Db    3725 AGGAGGGCGTTATGCATCTGCTCTTGAAGAAGAAGAACAACTAATCAGTCAGGCCAGCAG 3784

Qy    1083 CCGGGTTGCCAGCACCGCATCGGACGAAGCCCGTGAAAGTTCACGTAAATCGACCAGCCT 1142
             ||       |||||  | | | |    |||   || |     | |     |
Db    3785 TAAACAAGCAGAGGAAGCATCCCAAGTATCTAAAGAAGCATCCCAAGCGACAAATCAATT 3844

Qy    1143 GATTCAGGAAATGCTGAAAACAATGGAGAGCATTAACCAGTCGAAAGCATCCGCACTCGC 1202
           || ||  || |  |||| | ||| || ||||| ||||| || ||    || |||  |
Db    3845 AATACAAAAATTATTGAATATAATTGACAGCATCAACCAATCAAAGAATTCGGCAGCCAG 3904

Qy    1203 TGCTATCGCAGGCAATATTCGCGCTTAA 1230
           |   || || || || |||| ||||||
Db    3905 TCAGATTGCTGGTAACATTCGAGCTTAA 3932
```

```
RESULT   6
W10PPIBCA
LOCUS       W10PPIBCA     4761 bp     DNA              BCT       07-MAY-1996
DEFINITION  Plasmid pWR100 (from S.flexneri) antigen ippI, ipaB, ipaC genes,
            complete cds, and ippD gene, 3' end.
ACCESSION   M34849
NID         g155327
VERSION     M34849.1  GI:155327
KEYWORDS    antigen.
SOURCE      Plasmid pWR100 (clone: HB101/pHS[4108,5357].).
  ORGANISM  Plasmid pWR100
            plasmids.
REFERENCE   1  (bases 1 to 4761; 1 to 538; 3987 to 4761)
  AUTHORS   Baudry,B., Kaczorek,M. and Sansonetti,P.J.
  TITLE     Nucleotide sequence of the invasion plasmid antigen B and C genes
            (ipaB and ipaC) of Shigella flexneri
  JOURNAL   Microb. Pathog. 4 (5), 345-357 (1988)
  MEDLINE   89200844
COMMENT     Draft entry and computer-readable sequence for [1] kindly submitted
            by B.Baudry, 05-JUN-1990.
               Center for Vaccine Development
               University of Maryland at Baltimore
               10 South Pine Street
               Baltimore, MD 21201.
FEATURES             Location/Qualifiers
     source          1. .4761
                     /organism="Plasmid pWR100"
                     /plasmid="Plasmid pWR100"
                     /specific_host="M90T"
                     /db_xref="taxon:2661"
                     /clone="HB101/pHS[4108,5357]."
     CDS             <1. .549
                     /note="ORF"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA98422.1"
                     /db_xref="PID:g155328"
                     /db_xref="GI:155328"
                     /translation="KLVKKILSVFNLKQRFPQKNFGCPVNINKIRDSVIDKIKDSNSG
                     NQLFCWMSQERTTYVSSMINRSIDEMAIHNGVVLTSDNKKNIFAAIEKKFPDIKLDEK
                     SAQTSISHTALNEIASSGLRAKILKRYSSDMDLFNTQMKDLTNLVSSSVYDKIFNEST
                     KVLQIEISAEVLKAVYRQSNTN"
     -10_signal      556. .561
                     /gene="ippI"
     gene            556. .1073
                     /gene="ippI"
     RBS             595. .600
                     /gene="ippI"
     CDS             606. .1073
                     /gene="ippI"
                     /codon_start=1
                     /transl_table=11
                     /product="antigen"
                     /protein_id="AAA98423.1"
                     /db_xref="PID:g155329"
                     /db_xref="GI:155329"
                     /translation="MSLNITENESISTAVIDAINSGATLKDINAIPDDMMDDIYSYAY
                     DFYNKGRIEEAEVFFRFLCIYDFYNVDYIMGLAAIYQIKEQFQQAADLYAVAFALGKN
                     DYTPVFHTGQCQLRLKAPLKAKECFELVIQHSNDEKLKIKAQSYLDAIQDIKE"
     gene            1065. .2821
                     /gene="ipaB"
     RBS             1065. .1070
                     /gene="ipaB"
     CDS             1079. .2821
                     /gene="ipaB"
                     /codon_start=1
                     /transl_table=11
                     /product="antigen"
                     /protein_id="AAA98424.1"
                     /db_xref="PID:g155330"
                     /db_xref="GI:155330"
                     /translation="MHNVSTTTGFPLAKILTSTELGDNTIQAANDAANKLFSLTIAD
                     LTANQNINTTNAHSTSNILIPELKAPKSLNASSQLTLLIGNLIQILGEKSLIALTNKI
                     TAWKSQQQARQQKNLEFSDKINTLLSEIEGLTRDYEKQINKLKNADSKIKDLENKINQ
                     IQTRLSELDPESPEKKKLSREEIQLTIKKDAAVKDRTLIEQKTLSIHSKLIDKSMQLE
                     KEIDSFSAFSNTASAEQLSTQQKSLTGLASVTQLMATFIQLVGKNNEESLKNDLALFQ
                     SLQESRKTEMERKSDEYAAEVRKAEELNRVMGCVGKILGALLTIVSVVAAAFSGGASL
                     ALAAVGLALMVTDAIVQAATGNSFMEQALNPIMKAVIEPLIKLLSDAFTKMLEGLGVD
                     SKKAKMIGSILGAIAGALVLVAAVVLVATVGKQAAAKLAENIGKIIGKTLTDLIPKFL
                     KNFSSQLDDLITNAVARLNKFLGAAGDEVISKQIISTHLNQAVLLGESVNSATQAGGS
```

```
                     VASAVFQNSASTNLADLTLSKYQVEQLSKYISEAIEKFGQLQEVIADLLASMSNSQAN
                     RTDVAKAILQQTTA"
     gene            2723. .3932
                     /gene="ipaC"
     -35_signal      2723. .2728
                     /gene="ipaC"
     -10_signal      2745. .2750
                     /gene="ipaC"
     RBS             2829. .2834
                     /gene="ipaC"
     CDS             2841. .3932
                     /gene="ipaC"
                     /codon_start=1
                     /transl_table=11
                     /product="antigen"
                     /protein_id="AAA98425.1"
                     /db_xref="PID:g155331"
                     /db_xref="GI:155331"
                     /translation="MEIQNTKPTQTLYTDISTKQTQSSSETQKSQNYQQIAAHIPLNV
                     GKNPVLTTTLNDDQLLKLSEQVQHDSEIIARLTDKKMKDLSEMSHTLTPENTLDISSL
                     SSNAVSLIISVAVLLSALRTAETKLGSQLSLIAFDATKSAAENIVRQGLAALSSSITG
                     AVTQVGITGIGAKKTHSGISDQKGALRKNLATAQSLEKELAGSKLGLNKQIDTNITSP
                     QTNSSTKFLGKNKLAPDNISLSTEHKTSLSSPDISLQDKIDTQRRTYELNTLSAQQKQ
                     NIGRAIMETSAVAGNISTSGGRYASALEEEEQLISQASSKQAEEASQVSKEASQATNQ
                     LIQKLLNIIDSINQSKNSAASQIAGNIRA"
     RBS             3971. .3977
                     /gene="ipaD"
     gene            3971. .4761
                     /gene="ipaD"
     CDS             3983. .>4761
                     /gene="ipaD"
                     /codon_start=1
                     /transl_table=11
                     /product="antigen"
                     /protein_id="AAA98426.1"
                     /db_xref="PID:g552045"
                     /db_xref="GI:552045"
                     /translation="MNITTLTNSISTSSFSPNNTNGSSTETVNSDIKTTTSSHPVSSL
                     TMLNDILHNIRTTNQALKKELSQKTLTKTSLEEIALHSSQISMDVNKSAQLLDILSRN
                     EYPINKDARELLHSAPKEAELDGDQMISHRELWAKIANSINDINEQYLKVYEHAVSSY
                     TQMYQDFSAVLSSLAGWISPGGNDGNSVKLQVNSLKKALEELKEKYKDKPLYPANNTV
                     SQEQANKWLTELGGTIGKVSQKNGGYVVSINMTPIDNMLKS"
BASE COUNT     1769 a    952 c    805 g   1235 t
ORIGIN


  Query Match              5.4%;  Score 66.4;  DB 1;  Length 4761;
  Best Local Similarity   48.2%;  Pred. No. 1.3e-08;
  Matches  187;  Conservative    0;  Mismatches  201;  Indels    0;  Gaps    0;

Qy    843 TAATAAACAAATCTCCCCTGAACATCAGGCTATTCTGTCGAAACGTCTTGAGTCTGTCGA 902
          ||||| |    | ||  |||||||| |  ||  |||        |   |  ||| |  |
Db   3545 TAATATATCCCTGTCAACTGAACATAAAACTTCTCTTAGTTCTCCCGATATTTCTTTGCA 3604

Qy    903 ATCCGATATTCGTCTTGAGCAGAATACCATGGATATGACCCGAATCGATGCGCGCAAGAT 962
            | |||       ||   |  ||    ||  | |       |   |||||   | |
Db   3605 GGATAAAATTGACACCCAGAGAAGAACTTACGAGCTCAATACCCTTTCTGCGCAGCAAAA 3664

Qy    963 GCAGATGACGGGCGATCTGATTATGAAGAACTCGGTCACGGTCGGTGGTATTGCAGGGGC 1022
          || |  |  |||  |   |  ||| | |  || | |   |  |||  |||  |     |
Db   3665 ACAAAACATTGGCCGTGCAACAATGGAAACATCAGCCGTTGCTGGTAATATATCCACATC 3724

Qy   1023 GTCCGGGCAGTACGCCGCTACTCAGGAACGTTCCGAGCAGCAAATTAGCCAGGTGAATAA 1082
            ||||  || ||  || |||  |||      || || | ||| || ||||  |  |
Db   3725 AGGAGGGCGTTATGCATCTGCTCTTGAAGAAGAAGAACAACTAATCAGTCAGGCCAGCAG 3784

Qy   1083 CCGGGTTGCCAGCACCGCATCGGACGAAGCCCGTGAAAGTTCACGTAAATCGACCAGCCT 1142
              ||       |||||  | | | |    |||   || |       | |       |
Db   3785 TAAACAAGCAGAGGAAGCATCCCAAGTATCTAAAGAAGCATCCCAAGCGACAAATCAATT 3844

Qy   1143 GATTCAGGAAATGCTGAAAACAATGGAGAGCATTAACCAGTCGAAAGCATCCGCACTCGC 1202
          || ||  || |  |||| | ||| || ||||| ||||| || ||    || |||  |
Db   3845 AATACAAAAATTATTGAATATAATTGACAGCATCAACCAATCAAAGAATTCGGCAGCCAG 3904

Qy   1203 TGCTATCGCAGGCAATATTCGCGCTTAA 1230
          |   || || || || ||||| ||||||
Db   3905 TCAGATTGCTGGTAACATTCGAGCTTAA 3932


RESULT   7
SFPVR2
LOCUS       SFPVR2       7663 bp    DNA             BCT       12-SEP-1993
DEFINITION  Shigella flexineri plasmid YSH6000 virulence region-2 DNA
            containing 24k gene, 18k gene, ipaB gene, ipaC gene and ipaD gene.
ACCESSION   X15319
NID         g47052
VERSION     X15319.1  GI:47052
KEYWORDS    ipaB gene; ipaC gene; ipaD gene; virulence proteins.
SOURCE      Enterobacteriaceae.
  ORGANISM  Enterobacteriaceae
            Bacteria; Proteobacteria; gamma subdivision.
REFERENCE   1  (bases 1 to 7663)
  AUTHORS   Sasakawa,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAY-1989) Sasakawa C., Institute of Medical Science,
            University of Tokyo, Shirokanedai 4-6-1, Minato-ku Tokyo 108, Japan
REFERENCE   2  (bases 1010 to 6538)
  AUTHORS   Sasakawa,C., Adler,B., Tobe,T., Okada,N., Nagai,S., Komatsu,K. and
            Yoshikawa,M.
  TITLE     Functional organization and nucleotide sequence of virulence
            Region-2 on the large virulence plasmid in Shigella flexneri 2a
  JOURNAL   Mol. Microbiol. 3 (9), 1191-1201 (1989)
  MEDLINE   90014179
COMMENT     Region 2 has three promoters that transcribe RNAs of 4.0, 4.5 and
            7.5kb. The 4.0 kb transcript produces the 24,18kD, IpaB and IpaC
            proteins. The 4.5 kb transcript produces the ipaD protein. The
            7.5kb transcript covers the entire region 2.
            Data kindly reviewed (20-NOV-1989) by Sasakawa C.
FEATURES             Location/Qualifiers
     source          1. .7663
                     /organism="Enterobacteriaceae"
                     /plasmid="YSH6000"
                     /db_xref="taxon:543"
     promoter        1291. .1296
                     /note="-35 region for 4kb transcript"
     promoter        1312. .1318
                     /note="-10 region for 4kb transcript"
     RBS             1360. .1362
                     /note="put. ribosome binding site"
     CDS             1370. .1996
                     /note="24k gene (AA 1-208)"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA33379.1"
                     /db_xref="PID:g47053"
                     /db_xref="GI:47053"
                     /db_xref="SWISS-PROT:P33548"
                     /translation="MQILNKILPQVEFAIPRPSFDSLSRNKLVKKILSVFNLKQRFPQ
                     KNFGCPVNINKIRDSVIDKIKDSNSGNQLFCWMSQERTTYVSSMINRSIDEMAIHNGV
                     VLTSDNKKNIFAAIEKKFPDIKLDEKSAQTSISHTALNEIASSGLRAKILKRYSSDMD
                     LFNTQMKDLTNLVSSSVYDKIFNESTKVLQIEISAEVLKAVYRQSNTN"
     RBS             2043. .2045
                     /note="put. ribosome binding site"
     CDS             2053. .2520
                     /note="18k gene (AA 1-155)"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA33380.1"
                     /db_xref="PID:g47054"
                     /db_xref="GI:47054"
                     /db_xref="SWISS-PROT:P18008"
                     /translation="MSLNITENESISTAVIDAINSGATLKDINAIPDDMMDDIYSYAY
                     DFYNKGRIEEAEVFFRFLCIYDFYNVDYIMGLAAIYQIKEQFQQAADLYAVAFALGKN
                     DYTPVFHTGQCQLRLKAPLKAKECFELVIQHSNDEKLKIKAQSYLDAIQDIKE"
     RBS             2513. .2515
                     /note="put. ribosome binding site"
```

```
     CDS             2526. .4268
                     /note="ipaB gene (AA 1-580)"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA33381.1"
                     /db_xref="PID:g47055"
                     /db_xref="GI:47055"
                     /db_xref="SWISS-PROT:P18011"
                     /translation="MHNVSTTTTGFPLAKILASTELGDNTIQAANDAANKLFSLTIAD
                     LTANQNINTTNAHSTSNILIPELKAPKSLNASSQLTLLIGNLIQILGEKSLTALTNKI
                     IAWKSQQQARQQKNLEFSDKINTLLSETEGLTRDYEKQINKLKNADSKIKDLENKINQ
                     IQTRLSELDPESPEKKKLSREEIQLTIKKDAAVKDRTLIEQKTLSIHSKLTDKSMQLE
                     KEIDSFSAFSNTASAEQLSTQQKSLTGLASVTQLMATFIQLVGKNNEESLKNDLALFQ
                     SLQESRKTEMERKSDEYAAEVRKAEELNRVMGCVGKILGALLTIVSVVAAAFSGGASL
                     ALAAVGLALMVTDAIVQAATGNSFMEQALNPIMKAVIEPLIKLLSDAFTKMLEGLGVD
                     SKKAKMIGSILGAIAGALVLVAAVVLVATVGKQAAAKLAENIGKIIGKTLTDLIPKFL
                     KNFSSQLDDLITNAVARLNKFLGAAGDEVISKQIISTHLNQAVLLGESVNSATQAGGS
                     VASAVFQNSASINLADLTLSKYQVEQLSKYISEAIEKFGQLQEVIADLLASMSNSQAN
                     RTDVAKAILQQTTA"
     RBS             4224. .4226
                     /note="put. ribosome binding site"
     CDS             4231. .5379
                     /note="ipaC gene (AA 1-382)"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA33382.1"
                     /db_xref="PID:g47056"
                     /db_xref="GI:47056"
                     /db_xref="SWISS-PROT:P18012"
                     /translation="MLQKQFCNKLLLDTNKENVMEIQNTKPTQILYTDISTKQTQSSS
                     ETQKSQNYQQIAAHIPLNVGKNPVLTTTLNDDQLLKLSEQVQHDSEIIARLTDKKMKD
                     LSEMSHTLTPENTLDISSLSSNAVSLIISVAVLLSALRTAETKLGSQLSLIAFDATKS
                     AAENIVRQGLAALSSSITGAVTQVGITGIGAKKTHSGISDQKGALRKNLATAQSLEKE
                     LAGSKLGLNKQIDTNITSPQTNSSTKFLGKNKLAPDNISLSTEHKTSLSSPDISLQDK
                     IDTQRRTYELNTLSAQQKQNIGRATMETSAVAGNISTSGGRYASALEEEEQLISQASS
                     KQAEEASQVSKEASQATNQLIQKLLNIIDSINQSKNSTASQIAGNIRA"
     promoter        5376. .5381
                     /note="-35 region for 4.5kb transcript"
     promoter        5396. .5401
                     /note="-10 region for 4.5kb transcript"
     RBS             5421. .5423
                     /note="put. ribosome binding site"
     CDS             5430. .6428
                     /note="ipaD gene (AA 1-332)"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA33383.1"
                     /db_xref="PID:g47057"
                     /db_xref="GI:47057"
                     /db_xref="SWISS-PROT:P18013"
                     /translation="MNITTLTNSISTSSFSPNNTNGSSTETVNSDIKTTTSSHPVSSL
                     TMLNDTLHNIRTTNQALKKELSQKTLTKTSLEEIALHSSQISMDVNKSAQLLDILSRH
                     EYPINKDARELLHSAPKEAELDGDQMISHRELWAKIANSINDINEQYLKVYEHAVSSY
                     TQMYQDFSAVLSSLAGWISPGGNDGNSVKLQVNSLKKALEELKEKYKDKPLYPANNTV
                     SQEQANKWLTELGGTIGKVSQKNGGYVVSINMTPIDNMLKSLDNLGGNGEVVLDNAKY
                     QAWNAGFSAEDETMKNNLQTLVQKYSNANSIFDNLVKVLSSTISSCTDTDKLFLHF"
BASE COUNT     2845 a   1501 c   1246 g   2071 t
ORIGIN


  Query Match             5.3%;  Score 64.8;  DB 1;  Length 7663;
  Best Local Similarity   47.9%;  Pred. No. 3.9e-08;
  Matches  186;  Conservative    0;  Mismatches  202;  Indels    0;  Gaps    0;

Qy     843 TAATAAACAAATCTCCCCTGAACATCAGGCTATTCTGTCGAAACGTCTTGAGTCTGTCGA 902
           ||||| |    | ||  |||||||| |  ||  |||     |   |  ||| |  |
Db    4992 TAATATATCCCTGTCAACTGAACATAAAACTTCTCTTAGTTCTCCCGATATTTCTTTGCA 5051

Qy     903 ATCCGATATTCGTCTTGAGCAGAATACCATGGATATGACCCGAATCGATGCGCGCAAGAT 962
             | |||       ||   |  ||    ||  | |       |   |||||   | |
Db    5052 GGATAAAATTGACACCCAGAGAAGAACTTACGAGCTCAATACCCTTTCTGCGCAGCAAAA 5111

Qy     963 GCAGATGACGGGCGATCTGATTATGAAGAACTCGGTCACGGTCGGTGGTATTGCAGGGGC 1022
           || |  |  |||  |   |  ||| | |  || | |   |  |||  |||  |     |
Db    5112 ACAAAACATTGGCCGTGCAACAATGGAAACATCAGCCGTTGCTGGTAATATATCCACATC 5171

Qy    1023 GTCCGGGCAGTACGCCGCTACTCAGGAACGTTCCGAGCAGCAAATTAGCCAGGTGAATAA 1082
            ||||  || ||  || |||  |||       || || | ||| || ||||  |  |
Db    5172 AGGAGGGCGTTATGCATCTGCTCTTGAAGAAGAAGAACAACTAATCAGTCAGGCCAGCAG 5231

Qy    1083 CCGGGTTGCCAGCACCGCATCGGACGAAGCCCGTGAAAGTTCACGTAAATCGACCAGCCT 1142
               ||       |||||  | | | |    |||   || |       | |       |
Db    5232 TAAACAAGCAGAGGAAGCATCCCAAGTATCTAAAGAAGCATCCCAAGCGACAAATCAATT 5291

Qy    1143 GATTCAGGAAATGCTGAAAACAATGGAGAGCATTAACCAGTCGAAAGCATCCGCACTCGC 1202
            || ||  || |  |||| | ||| || ||||| ||||| || ||    ||  || |
Db    5292 AATACAAAAATTATTGAATATAATTGACAGCATCAACCAATCAAAGAATTCGACAGCCAG 5351

Qy    1203 TGCTATCGCAGGCAATATTCGCGCTTAA 1230
            |   || || || || |||||  ||||||
Db    5352 TCAGATTGCTGGTAACATTCGAGCTTAA 5379


RESULT   8
SDIPABCD
LOCUS       SDIPABCD     6869 bp    DNA             BCT       24-JUL-1991
DEFINITION  Shigella dysenteriae ipaBCD genes.
ACCESSION   X60777
NID         g46932
VERSION     X60777.1  GI:46932
KEYWORDS    ipaB gene; ipaC gene; ipaD gene; ippI gene.
SOURCE      Shigella dysenteriae.
  ORGANISM  Shigella dysenteriae
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Shigella.
REFERENCE   1  (bases 1 to 6869)
  AUTHORS   Palchaudhuri,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUN-1991) S. Palchaudhuri, Wayne State Univers.School
            of Medicine, Department of Immunology and, Microbiology, 540 E.
            Canfield, Detroit, MI 48201, USA
REFERENCE   2  (bases 1 to 6869)
  AUTHORS   Yao,R. and Palchaudhuri,S.
  TITLE     Nucleotide sequence of the ipaBCD structure genes of Shigella
            dysenteriae
  JOURNAL   Unpublished
FEATURES             Location/Qualifiers
     source          1. .6869
                     /organism="Shigella dysenteriae"
                     /strain="CG097"
                     /db_xref="taxon:622"
     CDS             317. .943
                     /note="24kD ORF"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA43188.1"
                     /db_xref="PID:g46933"
                     /db_xref="GI:46933"
                     /db_xref="SWISS-PROT:Q04314"
                     /translation="MQILNKILPQVEFAIPRPSFNSLSRNKLVKKILSVFNLKQRFPQ
                     KNFGCTVNINKIRDNVIDKIKDSNSGNQLFCWMSQERTTYVSSMINRSIDEMAIHNGV
                     VLTSDNKKNIFAAIEKKFPDIKLDEKSAQTSISHTALNEIASSGLRAKILKRYSSDMD
                     LFNTQMKDLTNLVSSSVYDKIFNESTKVLQIEISAEVLKAVYRQSNTN"
     CDS             1000. .1467
                     /note="IppI ORF"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA43189.1"
                     /db_xref="PID:g46934"
                     /db_xref="GI:46934"
                     /db_xref="SWISS-PROT:P18008"
                     /translation="MSLNITENESISTAVIDAINSGATLKDINAIPDDMMDDIYSYAY
                     DFYNKGRIEEAEVFFRFLCIYDFYNVDYIMGLAAIYQIKEQFQQAADLYAVAFALGKN
                     DYTPVFHTGQCQLRLKAPLKAKECFELVIQHSNDEKLKIKAQSYLDAIQDIKE"
     CDS             1473. .3215
                     /note="IpaB ORF"
```

```
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA43190.1"
                     /db_xref="PID:g46935"
                     /db_xref="GI:46935"
                     /db_xref="SWISS-PROT:Q03945"
                     /translation="MHNVNTTTTGLSLAKILASTELGDNTIQAGNDAANKLFSLTIAD
                     LTANKNINTTNAHSTSNILIPELKAPKSLNASSQLTLLIGNLIQILGEKSLTALTNKI
                     TAWKSQQQARQQKNLEFSDKINTLLSETEGLTRDYEKQINKLKNADSKIKDLENKINQ
                     IQTRLSELDPDSPEKKKLSREEIQLTIKKDAAVKDRTLIEQKTLSIHSKLTDKSMQLE
                     KEIDSFSAFSNTASAEQLSTQQKSLTGLASVTQLMATFIQLVGKNNEESLKNDLALFQ
                     SLQESRKTEMERKSDEYAAEVRKAEELNRVMGCVGKILGALLTIVSVVAAAFSGGASL
                     ALADVGLALMVTDAIVQAATGNSFMEQALNPIMKAVIEPLIKLLSDAFTKMLEGLGVD
                     SKKAKMIGSILGAIAGALVLVAAVVLVATVGKQAAAKLAENIGKIIGKTLTDLIPKFL
                     KNFSSQLDDLITNAVARLNKFLGAAGDEVISKQIISTHLNQAVLLGESVNSATQAGGS
                     VASAVFQNSASTNLADLTLSKYQVEQLSKYISEAIEKFGQLQEVIADLLASMSNSQAN
                     RTDVAKAILQQTTA"
     CDS             3178. .4326
                     /note="IpaC ORF"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA43191.1"
                     /db_xref="PID:g46936"
                     /db_xref="GI:46936"
                     /db_xref="SWISS-PROT:Q03946"
                     /translation="MLQKQFCNKLLLDTNKENVMEIQNTKSTQILYTDISTKQTQSSS
                     ETQKSQNYQQIAAHIPLNVGKNPVLTTTLNDDQLLKLSEQVQHDSEIIARLTDKKMKD
                     LSEMSHTLTPENTLDISSLSSNAVSLIISVAVLLSALRTAETKLGSQLSLIAFDATKS
                     AAENIVRQGLAALSSSITGAVTQVGITGIGAKKTHSGISDQKGALRKNLATAQSLEKE
                     LAGSKLGLNKQIDTNITSPQTNSSTKFLGKNKLAPDNISLSTEHKTSLSSPDISLQDK
                     IDTQRRTYELNTLSAQQKQNIGRATMETSAVAGNISTSGGRYASALEEEEQLISQASS
                     KQAEEASQVSKEASQATNQLIQKLLNIIDNINQSRSSTASQIAGNIRA"
     CDS             4377. .5375
                     /note="IpaD ORF"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA43192.1"
                     /db_xref="PID:g46937"
                     /db_xref="GI:46937"
                     /db_xref="SWISS-PROT:Q03947"
                     /translation="MNITTLTNSISTSSFSPNNTNGSSTETVNSDIKTTTSSHPVSSL
                     TMLNDTLHNIRTTNQALKKELSQKTLTKTSLEEIALHSSQISMDVNKSAQLLNILSKT
                     EYPINKDARELLHSAPKEAELDGYEMISHRELWAKIANSINDINEQYLKVYEHAVSSY
                     TQMYQEFSAVLSSLAGWISPGGNDGNSVKLQVKSLKDALTTLKKNYEDKPLYPATNTV
                     SEQEANKWLTELGGTIGTVSAKNGGYVVSINMTPIYNMLNRLDNLGGNGEVVLDNAKY
                     QAWNAGFSAEDETMKNNLQTLVQKYSNANSIFDNLVKVLSSTISSCTDTDKLFLHF"
BASE COUNT     2518 a   1369 c   1135 g   1847 t
ORIGIN


  Query Match             5.1%;  Score 63.2;  DB 1;  Length 6869;
  Best Local Similarity   47.7%;  Pred. No. 1.1e-07;
  Matches  185;  Conservative    0;  Mismatches  203;  Indels    0;  Gaps    0;

Qy     843 TAATAAACAAATCTCCCCTGAACATCAGGCTATTCTGTCGAAACGTCTTGAGTCTGTCGA 902
           ||||| |    | ||  |||||||| |  ||  |||       |    |   ||| | |
Db    3939 TAATATATCCCTGTCAACTGAACATAAAACTTCTCTTAGTTCTCCCGATATTTCTTTGCA 3998

Qy     903 ATCCGATATTCGTCTTGAGCAGAATACCATGGATATGACCCGAATCGATGCGCGCAAGAT 962
             | |||       ||   |  ||    ||  | |      |   |||||   | |
Db    3999 GGATAAAATTGACACCCAGAGAAGAACTTACGAGCTCAATACCCTTTCTGCGCAGCAAAA 4058

Qy     963 GCAGATGACGGGCGATCTGATTATGAAGAACTCGGTCACGGTCGGTGGTATTGCAGGGGC 1022
            || |  |  |||  |   |  ||| | |  || | |   |  ||| ||| |     |
Db    4059 ACAAAACATTGGCCGTGCAACAATGGAAACATCAGCCGTTGCTGGTAATATATCCACATC 4118

Qy    1023 GTCCGGGCAGTACGCCGCTACTCAGGAACGTTCCGAGCAGCAAATTAGCCAGGTGAATAA 1082
           ||||  || ||  ||  ||| |||       || || | ||| || ||||   |  |
Db    4119 AGGAGGGCGTTATGCATCTGCTCTTGAAGAAGAAGAACAACTAATCAGTCAGGCCAGCAG 4178

Qy    1083 CCGGGTTGCCAGCACCGCATCGGACGAAGCCCGTGAAAGTTCACGTAAATCGACCAGCCT 1142
             ||       |||||  | | | |    |||   || |      | |      |
Db    4179 TAAACAAGCAGAGGAAGCATCCCAAGTATCTAAAGAAGCATCCCAAGCGACAAATCAATT 4238

Qy    1143 GATTCAGGAAATGCTGAAAACAATGGAGAGCATTAACCAGTCGAAAGCATCCGCACTCGC 1202
           || ||  || | ||||| | ||| || | ||| ||||| || |     ||  ||  |
Db    4239 AATACAAAAATTACTGAATATAATTGACAACATCAACCAATCAAGGAGTTCTACAGCCAG 4298

Qy    1203 TGCTATCGCAGGCAATATTCGCGCTTAA 1230
           |   || || || || ||||| ||||||
Db    4299 TCAGATTGCTGGTAACATTCGAGCTTAA 4326


RESULT   9
I66494/c
LOCUS       I66494       7218 bp    DNA             PAT       23-DEC-1997
DEFINITION  Sequence 14 from patent US 5670367.
ACCESSION   I66494
NID         g2724471
VERSION     I66494.1  GI:2724471
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 7218)
  AUTHORS   Dorner,F., Scheiflinger,F. and Falkner,F.Gunter.
  TITLE     Recombinant fowlpox virus
  JOURNAL   Patent: US 5670367-A 14 23-SEP-1997;
FEATURES             Location/Qualifiers
     source          1. .7218
                     /organism="unknown"
BASE COUNT     1944 a   1491 c   1486 g   1929 t    368 others
ORIGIN


  Query Match             4.1%;  Score 50.8;  DB 5;  Length 7218;
  Best Local Similarity   4.0%;  Pred. No. 0.00044;
  Matches   16;  Conservative  219;  Mismatches  161;  Indels    0;  Gaps    0;

Qy     479 AAACGACGGCAAGCTCCATGATGCGCGAAGGGATGAATGCGTTGTCCGGTAGTATTTCCC 538
           ::: :: :: :::    : :: : : ::::::: ::: : :  :   :: :: :
Db    1434 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1375

Qy     539 AGAGCGCGCTTCAGTTGGGGATCACTGGCGTGGGCGCCAAACTGGAATATAAGGGGCTGC 598
           :::: : :    ::  :::::  :  :: : ::: :  :::  :::: : ::::::  :
Db    1374 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1315

Qy     599 AGAATGAAAGAGGCGCGCTTAAACATAATGCCGCGAAGATCGATAAACTGACCACTGAAA 658
           :::: ::::::::: : :   ::: : :: :  : :::::  :: :::  ::  :  ::::
Db    1314 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1255

Qy     659 GCCACAGTATTAAAAACGTGCTGAACGGGCAGAATAGCGTCAAACTCGGTGCTGAAGGCG 718
           :  : :: :  ::::: : :  ::: ::: :::: :: :  :::   :: :  ::::: :
Db    1254 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1195

Qy     719 TCGATTCTCTGAAATCGTTAAATATGAAGAAAACCGGTACCGATGCGACGAAAAATCTTA 778
             ::      ::::  :  ::: : ::::::::  :: :  :: : :: ::::::    :
Db    1194 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1135

Qy     779 ATGATGCGACGCTTAAATCTAATGCCGGAACCAGCGCCACGGAAAGTCTGGGTATTAAAG 838
           : :: : :: :   :::   :: :  :::: :: :  : ::::::   ::: :  ::::
Db    1134 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1075

Qy     839 ACAGTAATAAACAAATCTCCCCTGAACATCAGGCTA 874
           : :: :: |    || ||||  || |  ||| | ||
Db    1074 RRRRRRRRATCGCAAGCTCCCTCGACCTGCAGCCAA 1039


RESULT  10
AC004836/c
LOCUS       AC004836   103694 bp    DNA             PRI       28-APR-1999
DEFINITION  Homo sapiens PAC clone DJ0568B10 from 7q31.1-q31.2, complete
            sequence.
ACCESSION   AC004836
NID         g4508152
VERSION     AC004836.2  GI:4508152
```

KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 103694)
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
REFERENCE 2 (bases 1 to 103694)
AUTHORS Bauer,C. and Biewald,T.
TITLE The sequence of Homo sapiens PAC clone DJ0568B10
JOURNAL Unpublished (1999)
REFERENCE 3 (bases 1 to 103694)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (12-JUN-1998) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 103694)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (24-MAR-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 5 (bases 1 to 103694)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (28-APR-1999) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Mar 24, 1999 this sequence version replaced gi:3213161.
SUBMITTED BY: WUGSC
Genome Sequencing Center
Department of Genetics
Washington University
St. Louis MO 63108, USA
http://genome.wustl.edu/gsc
mailto:sapiens@watson.wustl.edu

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see http://www.nhgri.nih.gov/DIR/GTB/CHR7 , send mailto:egreen@nhgri.nih.gov , or see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
This clone was derived from human PAC library RPCI-4, prepared by Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://bacpac.med.buffalo.edu) using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor.
The clone may be obtained either from Genome Systems, Inc. (http://www.genomesystems.com) or Research Genetics, Inc. (http://www.resgen.com); or from Pieter de Jong.
VECTOR: pCYPAC2

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is DJ1106G10, 200 bp overlap; the clone sequenced to the right is RG332P12, 200 bp overlap. Actual start of this clone is at base position 197 of DJ0568B10; actual end is at 88190 of RG332P12.

FEATURES Location/Qualifiers
source 1. .103694
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q31.1-q31.2"
/clone_lib="RPCI-4"
/clone="DJ0568B10"
repeat_region 1. .44
/rpt_family="Alu"
misc_feature 531. .770
/note="match to EST H91012 (NID:g1081442) yu89d02.r1"
repeat_region 1064. .1189
/rpt_family="MIR"
repeat_region 2043. .2164
/rpt_family="MIR"
repeat_region 2211. .2571
/rpt_family="MaLR"
repeat_region 2572. .2861
/rpt_family="Alu"
repeat_region 2862. .2921
/rpt_family="MaLR"
repeat_region 4779. .4900
/rpt_family="MER1_type"
repeat_region 5102. .5320
/rpt_family="MIR"
repeat_region 5330. .5758
/rpt_family="L1"
repeat_region 5870. .5998
/rpt_family="Alu"
repeat_region 6051. .6124
/rpt_family="MIR"
repeat_region 6818. .7062
/rpt_family="MIR"
repeat_region 7388. .7599
/rpt_family="MIR"
repeat_region 7940. .8160
/rpt_family="MIR"
repeat_region 8507. .8626
/rpt_family="MIR"
repeat_region 9149. .9170
/rpt_family="AT_rich"
repeat_region 9296. .9319
/rpt_family="AT_rich"
repeat_region 9323. .9394
/rpt_family="L1"
repeat_region 9473. .9521
/rpt_family="POLY_A"
repeat_region 9522. .9578
/rpt_family="L1"
misc_feature 9530. .9802
/note="match to EST H81829 (NID:g1059918) ys68f09.s1"
gene <9855. .>51335
/gene="WUGSC:H_DJ0568B10.1"
CDS join(<9855. .10020,13079. .13176,23638. .23732,25137. .25241, 30349. .30497,33385. .33702,44014. .44069,49260. .49431, 51085. .>51335)
/gene="WUGSC:H_DJ0568B10.1"
/note="similar to cadherin and Drosophila Fat protein; similar to CAA60685 (PID:g1107687); H_DJ0568B10.1"
/codon_start=1
/evidence=not_experimental
/protein_id="AAD28068.1"
/db_xref="PID:g4699969"
/db_xref="GI:4699969"
/translation="VVTTGMEQLDFETGPNIFDLQIYVKDEVGVTDLQVLTVQVTDVN EPPQFQGNLAEGLHLYIVERANPGFIYQVEAFDPEDTSRNIPLSYFLISPPKSFRMSA NGTLFSTTELDFEAGHRSFHLIVEVRDSGGLKASIELQVNIVNLNDEVPRFTSPTRVY

```
                    TVLEELSPGTIVANITAEDPDDEGFPSHLLYSITTVSKYFMINQLTGTIQVAQRIDRD
                    AGELRQNPTISLEVLVKDRPYGGQENRIQITFIVEDVNDNPATCQKFTFRYAHFESWA
                    GHSISLVALCPSASLVYSQENYKHLEKQLSRINRYYENPSNLAAGNKYTVIIQVTVNI
                    LEENDEKPICTPNSYFLALPVDLKVGTNIQNFKLTCTDLDSSPRSFRYSIGPGNVNNH
                    FTFSPNAGSNVTRLLLTSRFDYAGGFDKIWDYKLLVYVTDDNLMSDRKKAEALVETGT
                    VTLSIKVIPHPTTIIITTTPR"
     misc_feature   9881. .10020
                    /gene="WUGSC:H_DJ0568B10.1"
                    /note="match to EST H39479 (NID:g915175)"
     repeat_region  10314. .10411
                    /rpt_family="L2"
     repeat_region  10466. .11073
                    /rpt_family="L2"
     repeat_region  11074. .11385
                    /rpt_family="Alu"
     repeat_region  11386. .11587
                    /rpt_family="L2"
     repeat_region  11588. .11890
                    /rpt_family="Alu"
     repeat_region  11891. .12160
                    /rpt_family="L2"
     repeat_region  12178. .12397
                    /rpt_family="MIR"
     misc_feature   13079. .13176
                    /gene="WUGSC:H_DJ0568B10.1"
                    /note="match to EST H39479 (NID:g915175)"
     misc_feature   13217. .13525
                    /gene="WUGSC:H_DJ0568B10.1"
                    /note="match to EST T52032 (NID:g653892) yb29a06.r1"
     repeat_region  13306. .13542
                    /rpt_family="MIR"
     misc_feature   13529. .13944
                    /gene="WUGSC:H_DJ0568B10.1"
                    /note="match to EST T51952 (NID:g653812) yb29a06.s1"
     repeat_region  13680. .13963
                    /rpt_family="Alu"
     misc_feature   13990. .14247
                    /gene="WUGSC:H_DJ0568B10.1"
                    /note="match to EST AA514828 (NID:g2254428) nh71h12.s1"
     misc_feature   14126. .14546
                    /gene="WUGSC:H_DJ0568B10.1"
                    /note="match to EST AA996205 (NID:g3182694) os01h03.s1"
     repeat_region  14560. .14591
                    /rpt_family="(TGAA)n"
     repeat_region  14621. .14675
                    /rpt_family="L2"
     repeat_region  14700. .14975
                    /rpt_family="MaLR"

  Query Match              3.3%;  Score 41;  DB 42;  Length 103694;
  Best Local Similarity   56.2%;  Pred. No. 0.44;
  Matches   77;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy     528 TAGTATTTCCCAGAGCGCGCTTCAGTTGGGGATCACTGGCGTGGGCGCCAAACTGGAATA 587
           ||| ||||  || |  |  ||||||    || || |||| |  ||| |    |  || |
Db   33189 TAGGATTTTTCACACTGTACTTCAGAGTTGGGTCCCTGGTGGCGGCCCTTGCCCAGAGTG 33130

Qy     588 TAAGGGGCTGCAGAATGAAAGAGGCGCGCTTAAACATAATGCCGCGAAGATCGATAAACT 647
           |  || |||  |||| |  ||  |  |||| | |   | | | |    || |   || |
Db   33129 GACTGGACTGGTGAATAACTGACTCCAGCTTGACCCAGAGGGCTCTCTTATTGGAGAATT 33070

Qy     648 GACCACTGAAAGCCACA 664
           |  |||||||| ||||
Db   33069 TATAACTGAAAGGCACA 33053


RESULT  11
AC005069
LOCUS       AC005069   204597 bp    DNA             PRI       05-MAY-1999
DEFINITION  Homo sapiens clone RG140B11, complete sequence.
ACCESSION   AC005069
NID         g4753224
VERSION     AC005069.2  GI:4753224
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 204597)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 204597)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-1998) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   3  (bases 1 to 204597)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAY-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On May 5, 1999 this sequence version replaced gi:3212922.
FEATURES             Location/Qualifiers
     source          1. .204597
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RG140B11"
BASE COUNT    58283 a  43152 c  42964 g  60198 t
ORIGIN


  Query Match              3.3%;  Score 41;  DB 42;  Length 204597;
  Best Local Similarity   56.2%;  Pred. No. 0.49;
  Matches   77;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy     528 TAGTATTTCCCAGAGCGCGCTTCAGTTGGGGATCACTGGCGTGGGCGCCAAACTGGAATA 587
           ||| ||||  || |  |  ||||||    || || |||| |  ||| |    |  || |
Db  135173 TAGGATTTTTCACACTGTACTTCAGAGTTGGGTCCCTGGTGGCGGCCCTTGCCCAGAGTG 135232

Qy     588 TAAGGGGCTGCAGAATGAAAGAGGCGCGCTTAAACATAATGCCGCGAAGATCGATAAACT 647
           |  || |||  |||| |  ||  |  |||| | |   | | | |    || |   || |
Db  135233 GACTGGACTGGTGAATAACTGACTCCAGCTTGACCCAGAGGGCTCTCTTATTGGAGAATT 135292

Qy     648 GACCACTGAAAGCCACA 664
           |  |||||||| ||||
Db  135293 TATAACTGAAAGGCACA 135309


RESULT  12
AC006424
LOCUS       AC006424   119660 bp    DNA             HTG       24-APR-1999
DEFINITION  Arabidopsis thaliania chromosome I clone F9L11, WORKING DRAFT
            SEQUENCE, 5 unordered pieces.
ACCESSION   AC006424
NID         g4678188
VERSION     AC006424.2  GI:4678188
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Arabidopsis thaliania.
  ORGANISM  Arabidopsis thaliania
            Unclassified.
REFERENCE   1  (bases 1 to 119660)
  AUTHORS   Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
            Altafi,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P.,
            Gonzalez,A., Kremenetskaia,I., Kim,C., Lenz,C., Li,J., Liu,S.,
            Luros,S., Schwartz,J., Shinn,P., Toriumi,M., Vysotskaia,V.S.,
            Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 119660)
  AUTHORS   Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
            Altafi,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P.,
            Gonzalez,A., Kremenetskaia,I., Kim,C., Lenz,C., Li,J., Liu,S.,
            Luros,S., Schwartz,J., Shinn,P., Toriumi,M., Vyotskaia,V.,
            Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JAN-1999) DNA Sequencing and Technology Center,
            Stanford University, 855 California Avenue, Palo Alto, CA 94304,
            USA
COMMENT     On Apr 24, 1999 this sequence version replaced gi:4165338.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     3638: contig of 3638 bp in length
            *     3639     3688: gap of unknown length
            *     3689     9307: contig of 5619 bp in length
            *     9308     9357: gap of unknown length
            *     9358    20920: contig of 11563 bp in length
            *    20921    20970: gap of unknown length
            *    20971    42198: contig of 21228 bp in length
            *    42199    42248: gap of unknown length
            *    42249   119660: contig of 77412 bp in length.
FEATURES             Location/Qualifiers
     source          1. .119660
                     /organism="Arabidopsis thaliania"
                     /clone="F9L11"
                     /chromosome="I"
BASE COUNT    38675 a  21361 c  21137 g  38273 t    214 others
ORIGIN


  Query Match              3.3%;  Score 40.2;  DB 35;  Length 119660;
  Best Local Similarity   48.9%;  Pred. No. 0.77;
  Matches  108;  Conservative    0;  Mismatches  113;  Indels    0;  Gaps    0;

Qy     624 TAATGCCGCGAAGATCGATAAACTGACCACTGAAAGCCACAGTATTAAAAACGTGCTGAA 683
           |||| |||  |||| |  |   |    |  ||||| |||  |||   |  | ||| |  |
Db  105917 TAATCCCGAAAAGAACACTCTCCAACACGTTGAAACCCATTGTAACCATGAAGTGTTTGA 105976

Qy     684 CGGGCAGAATAGCGTCAAACTCGGTGCTGAAGGCGTCGATTCTCTGAAATCGTTAAATAT 743
           ||  |||||   || ||  ||  ||  ||    || ||   ||| || |       |||
Db  105977 CGATGAGAATCTAGTTAACATCTTTGTAGACCATGTAGAGGATCTTAAGTTTGATGTTAT 106036

Qy     744 GAAGAAAACCGGTACCGATGCGACGAAAAATCTTAATGATGCGACGCTTAAATCTAATGC 803
           ||||  |||   | | |   |  | | |   |  |  ||   || || |||| | |
Db  106037 GAAGCCAACATATGCTGCAACATCTATACGCCCAACAGAACAGAAGCATAAACCCACGAG 106096

Qy     804 CGGAACCAGCGCCACGGAAAGTCTGGGTATTAAAGACAGTA 844
           |  |   |  | | |   ||    | |  | |||  || |
Db  106097 CACAGAAACATCGATGTCTAGAAAAGATCATCAAGTGAGGA 106137


RESULT  13
HUAC004020
LOCUS       HUAC004020 153773 bp    DNA             PRI       19-OCT-1998
DEFINITION  Homo sapiens Chromosome 16 BAC clone CIT987SK-A-319E8, complete
            sequence.
ACCESSION   AC004020
NID         g3219330
VERSION     AC004020.1  GI:3219330
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 153773)
  AUTHORS   Adams,M.D., Loftus,B.J., Zhou,L., Crosby,M., Fuhrmann,J.,
            Mason,T.M., Brandon,R., Kim,U.J., Kerlavage,A.R. and Venter,J.C.
  TITLE     Homo sapiens Chromosome 16 BAC clone CIT987SK-A-319E8
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 153773)
  AUTHORS   Adams,M.D. and Loftus,B.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JAN-1998) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, Email:
            bjloftus@tigr.org
REFERENCE   3  (bases 1 to 153773)
  AUTHORS   Adams,M.D. and Loftus,B.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUN-1998) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
COMMENT     On Jun 13, 1998 this sequence version replaced gi:3212145.
            Address all correspondence to: Mark Adams The Institute for Genomic
            Research 9712 Medical Center Dr, Rockville, MD 20850, USA e-mail
            address: mdadams@tigr.org.   The orientation of the sequence is
            from SP6 end to T7 end. Genes were identified by a combination of
            five methods including: XGRAIL (available by anonymous ftp from
            arthur.epm.ornl.gov), Genefinder (Phil Green, University of
            Washington), Genscan  (Chris Burge,
            http://gnomic.stanford.Edu/~chris/GENSCANW.html)searches of the
            complete sequence against a peptide database,and the Human gene
            Index database at TIGR (http://www.tigr.org/tdb/hgi/hgi.html).
            Genes without pepetide homolgy having spliced EST hits are termed
            'Unknown gene product'. Genes encoding tRNAs are predicted by
            tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
FEATURES             Location/Qualifiers
     source          1. .153773
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /map="16p13.1, 2q11.2"
                     /clone="A-319E8"
     STS             45251. .45382
                     /note="32826, A002D07, Chr. -, Homo sapiens"
                     /db_xref="dbSTS:G19948"
     STS             101253. .101392
                     /note="43400, SHGC-16460, Chr. 2, Homo sapiens"
                     /db_xref="dbSTS:G29219"
     mRNA            complement(join(<101402. .101642,103398. .103485,
                     104095. .104296,104829. .104917,105017. .105128,
                     105284. .105432,106264. .>106402))
                     /gene="A-319E8.1"
     gene            complement(<101402. .>106402)
                     /gene="A-319E8.1"
     CDS             complement(join(101402. .101642,103398. .103485,
                     104095. .104296,104829. .104917,105017. .105128,
                     105284. .105432,106264. .106402))
                     /gene="A-319E8.1"
                     /codon_start=1
                     /product="Unknown gene product"
                     /protein_id="AAC23493.1"
                     /db_xref="PID:g3219331"
                     /db_xref="GI:3219331"
                     /translation="MKDSLVLLGRVPAHPDSRCWFLAWNPAGTLLASCGGDRRIRIWG
                     TEGDSWICKSVLSEGHQRTVRKVAWSPCGNYLASASFDATTCIWKKNQDDFECVTTLE
                     GHENEVKSVAWAPSGNLLATCSRDKSVWVWEVDEEDEYECVSVLNSHTQDVKHVVWHP
                     SQELLASASYDDTVKLYREEEDDWVCCATLEGHESTVWSLAFDPSGQRLASCSDDRTV
                     RIWRQYLPGNEQGVACSGSDPSWKCICTLSGFHSRTIYDIAWCQLTGALATACGDDAI
                     RVFQEDPNSDPQQPTFSLTAHLHQAHSQDVNCVAWNPKEPGLLASCSDDGEVAFWKYQ
                     RPEGL"
     mRNA            join(<107372. .107615,117756. .117920,118638. .>118945)
                     /gene="A-319E8.2"
     gene            <107372. .>118945
                     /gene="A-319E8.2"
     CDS             join(107372. .107615,117756. .117920,118638. .118945)
                     /gene="A-319E8.2"
                     /codon_start=1
                     /product="Unknown gene product"
                     /protein_id="AAC23494.1"
                     /db_xref="PID:g3219332"
                     /db_xref="GI:3219332"
                     /translation="MYAPGGAGLPGGRRRRSPGGSALPKQPERSLASALPGALSITAL
                     CTALAEPAWLHIHGGTCSRQELGVSDVLGYVHPDLLKDFCMNPQTVLLLRVIAAFCFL
                     GILCSLSAFLLDVFGPKHPALKITRRYAFAHILTVLQCATVIGFSYWASELILAQQQQ
                     HKKYHGSQVYVTFAVSFYLVAGAGGASILATAANLLRHYPTEEEEQALELLSEMEENE
                     PYPAEYEVINQFQPPPAYTP"
```

```
BASE COUNT    42834 a  35082 c  33404 g  42450 t      3 others
ORIGIN


  Query Match             3.2%;  Score 38.8;  DB 11;  Length 153773;
  Best Local Similarity   53.2%;  Pred. No. 2;
  Matches   82;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy     148 GGGTTGAGTCCTACACTGAGCGCGCCTGCGCCAGGGGTATTAACGCAAACCCCCGGAACG 207
           ||| |||| || ||  |   | | ||  |||||| | ||| |  | |||     |  |
Db   96699 GGGCTGAGGACTGCAACGGCAGAGACTTTTCCAGGGCTTTTATCAGATACCAAGTCATAG 96758

Qy     208 ATCACGTCCTTTTTAAAAGCCAGTATTCAAAATACCGACATGAATCAGGATTTGAATGCT 267
           | || ||    ||||| |||||    || |  ||    |    |||   |   | ||||
Db   96759 AGCAGGTGGAATTTAAGAGCCAGAGAACAGAGGACACCAAGTGGTCACAGTGACATTGCT 96818

Qy     268 CTGGCAAATAATGTCACGACTAAAGCGAATGAGG 301
           |||  || |  | || | |  |||  | |   ||
Db   96819 CTGAAAACTTTTCTCTCCATGAAATGGCACAGGG 96852


RESULT  14
HSDGCR6
LOCUS       HSDGCR6      1080 bp    mRNA            PRI       03-MAY-1996
DEFINITION  H.sapiens mRNA for DGCR6 protein.
ACCESSION   X96484
NID         g1223740
VERSION     X96484.1  GI:1223740
KEYWORDS    DGCR6 gene.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1080)
  AUTHORS   Demczuk,S., Thomas,G. and Aurias,A.
  TITLE     Isolation of a novel gene from the DiGeorge syndrome critical
            region with homology to Drosophila gdl and to human LAMC1 genes
  JOURNAL   Hum. Mol. Genet. 5 (5), 633-638 (1996)
  MEDLINE   96311558
REFERENCE   2  (bases 1 to 1080)
  AUTHORS   Demczuk,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-1996) S. Demczuk, Institut Curie, Section de
            Recherche, INSERM U434 Genetique des Tumeurs, 26 rue dUlm, Paris,
            75231, France
FEATURES             Location/Qualifiers
     source          1. .1080
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="22"
                     /cell_line="HeLa"
                     /dev_stage="adult"
                     /map="q11.2"
                     /note="DiGeorge syndrome critical region"
     CDS             423. .677
                     /codon_start=1
                     /product="DGCR6"
                     /protein_id="CAA65339.1"
                     /db_xref="PID:e228435"
                     /db_xref="PID:g1223741"
                     /db_xref="GI:1223741"
                     /db_xref="SPTREMBL:Q14129"
                     /translation="MDQKIVLELDRKVADQQSTLEKAGVAGFYVTTNPQELMLQMNLL
                     ELIRKLQQRGCWAGKAALGLGGPWQLPAAQCDQKGSPVPP"
BASE COUNT      230 a    334 c    342 g    174 t
ORIGIN


  Query Match             3.1%;  Score 37.8;  DB 10;  Length 1080;
  Best Local Similarity   46.4%;  Pred. No. 2;
  Matches  123;  Conservative    0;  Mismatches  142;  Indels    0;  Gaps    0;

Qy     850 CAAATCTCCCCTGAACATCAGGCTATTCTGTCGAAACGTCTTGAGTCTGTCGAATCCGAT 909
           || | |   |||| | ||||||   || |  |  |   || ||| | || ||   |
Db     342 CATAACCTGCCTGTGCTTCAGGCGGCTCAGCAGCGAGAACTAGAGGCGGTGGAGCACCGG 401

Qy     910 ATTCGTCTTGAGCAGAATACCATGGATATGACCCGAATCGATGCGCGCAAGATGCAGATG 969
           || |||   ||||||    | |||||   ||      ||   | ||   |   | || ||
Db     402 ATCCGTGAGGAGCAGCGGGCGATGGACCAGAAGATCGTCCTGGAGCTGGACCGGAAGGTG 461

Qy     970 ACGGGCGATCTGATTATGAAGAACTCGGTCACGGTCGGTGGTATTGCAGGGGCGTCCGGG 1029
           | | | | | ||  |    | |   ||    ||| | |||  |    | | |  ||
Db     462 GCTGACCAGCAGAGCACACTGGAGAAGGCGGGGGTGGCTGGCTTCTACGTGACCACCAAC 521

Qy    1030 CAGTACGCCGCTACTCAGGAACGTTCCGAGCAGCAAATTAGCCAGGTGAATAACCGGGTT 1089
           |   | |      |  | | |      |  || | || | | || |  |    | | |
Db     522 CCACAGGAGCTGATGCTGCAGATGAACCTGCTGGAACTCATCCGGAAGCTGCAGCAGAGG 581

Qy    1090 GCCAGCACCGCATCGGACGAAGCCC 1114
           | | ||   |||  | | | ||||
Db     582 GGCTGCTGGGCAGGGAAGGCAGCCC 606


RESULT  15
LM5UTRDGN
LOCUS       LM5UTRDGN      462 bp    mRNA            INV       01-MAY-1997
DEFINITION  L.major mRNA, D gene, 5'UTR (clone Lm DRS-18)).
ACCESSION   Y09088
NID         g2065196
VERSION     Y09088.1  GI:2065196
KEYWORDS    5' UTR; D gene.
SOURCE      Leishmania major.
  ORGANISM  Leishmania major
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Leishmania.
REFERENCE   1  (bases 1 to 462)
  AUTHORS   Coulson,R.M.R.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 462)
  AUTHORS   Coulson,R.M.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-OCT-1996) R.M.R. Coulson, University of Cambridge,
            Pathology, Tennis Court Road, Cambridge, CB2 1QP, UK
FEATURES             Location/Qualifiers
     source          1. .462
                     /organism="Leishmania major"
                     /strain="LV39"
                     /specific_host="Sandfly"
                     /db_xref="taxon:5664"
                     /dev_stage="promastigote"
                     /clone="Lm DRS-18"
     gene            1. .462
                     /gene="D"
     5'UTR           1. .462
                     /gene="D"
BASE COUNT      129 a    145 c    113 g     75 t
ORIGIN


  Query Match             3.0%;  Score 37.2;  DB 36;  Length 462;
  Best Local Similarity   59.4%;  Pred. No. 2.6;
  Matches   63;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy     239 ATACCGACATGAATCAGGATTTGAATGCTCTGGCAAATAATGTCACGACTAAAGCGAATG 298
           | ||| | ||| | |||||    || |  | |   || || | | || |  | |||||
Db     146 AGACCAAGATGGACCAGGAGACAAAGGACCAGATGAAAAACGCCGCGGCGGAGGCGAAGG 205

Qy     299 AGGTTGTGCAAACCCAGTTACGCGAGCAGCAGGCAGAAGTCGGAAA 344
           |   ||||||   | || | |  |||| | |||  || || || ||
Db     206 ACAATGTGCACGACAAGATCCAGGAGCTGAAGGACGATGTGGGGAA 251


Search completed: October  5, 1999, 17:58:36
Job time: 11359 sec
```



This Page Blank (uspto)

OM of: US-09-068-804-5 to: EST:* out_format : pfs

Date: Oct 5, 1999 7:34 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
-MODEL=frame-_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09068804/runat_05101999_145134_10306/app_query.fasta.1
-DB=EST -QFMT=fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=pam150 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=escore -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfs -NORM=stat -USER=US09068804 -NCPU=6 -ICPU=3 -WAIT
-THREADS=1

Search information block:
Query: US-09-068-804-5
Query length: 290
Database: EST:*
Database sequences: 2546578
Database length: 986266752
Search time (sec): 940.790000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ..
gb_est9:AA107012	+	90.50	200.72	0.0115	377	! AA107012 ml93g04.r1 Stratagene
gb_est17:AA615768	+	88.00	191.99	0.0353	478	! AA615768 vo72f06.r1 Barstead mc
gb_est18:AA697508	-	88.00	190.22	0.0442	558	! AA697508 HL02674.5prime HL Dros
gb_est18:AA697063	-	86.00	184.09	0.0971	626	! AA697063 GM09027.5prime GM Dros
gb_est16:AA541225	-	85.00	181.92	0.1283	613	! AA541225 LD21073.5prime LD Dros
gb_est20:AA816702	-	83.00	180.10	0.1620	471	! AA816702 LD04266.5prime LD Dros
gb_est19:AA539284	-	83.00	179.79	0.1685	484	! AA539284 LD18968.5prime LD Dros
gb_est11:AA264493	-	84.00	179.09	0.1844	636	! AA264493 LD07632.5prime LD Dros
gb_est20:AA817581	-	83.00	178.17	0.2075	558	! AA817581 LD23545.5prime LD Dros
gb_est19:AA540277	-	83.00	178.05	0.2108	564	! AA540277 LD19717.5prime LD Dros
gb_est14:AA392365	-	83.00	177.77	0.2184	578	! AA392365 LD11385.5prime LD Dros
gb_est19:AA438684	-	83.00	176.75	0.2489	632	! AA438684 LD13011.5prime LD Dros
gb_est21:AA978975	-	83.00	175.62	0.2878	698	! AA978975 LD33203.5prime LD Dros
gb_est21:AA979658	-	83.00	174.95	0.3134	740	! AA979658 LD34364.5prime LD Dros
gb_est21:AA979513	-	83.00	174.76	0.3215	753	! AA979513 LD34126.5prime LD Dros
gb_est28:AI512595	-	82.00	173.66	0.3701	671	! AI512595 LD44514.5prime LD Dros
gb_est17:AA658052	+	81.00	172.48	0.4302	602	! AA658052 nr83g05.s1 NCI_CGAP_Pr
gb_est18:AA694997	-	80.50	172.29	0.4412	551	! AA694997 GM01919.5prime GM Dros
gb_est16:AA540452	-	77.00	165.93	0.9970	459	! AA540452 LD19946.5prime LD Dros
gb_est20:AA898980	-	77.50	165.52	1.05	529	! AA898980 NCM4D10T3 Mycelial Neu
gb_est19:AA201134	-	78.00	165.50	1.05	589	! AA201134 LD03806.5prime LD Dros
gb_est18:AA695555	-	77.50	165.48	1.06	531	! AA695555 GM03025.5prime GM Dros
gb_est21:AA933089	+	78.00	165.20	1.10	605	! AA933089 om85g05.s1 NCI_CGAP_Ki
gb_est19:AA390945	-	77.00	165.00	1.12	498	! AA390945 LD11786.5prime LD Dros
gb_est16:AA541064	-	77.00	163.79	1.31	554	! AA541064 LD20851.5prime LD Dros
gb_est32:AI722732	+	76.50	163.55	1.35	509	! AI722732 fc30h04.y1 Zebrafish W
gb_est4:D22789	+	77.50	162.56	1.54	686	! D22789 RICC1332A Rice callus Or
gb_est18:AA694773	-	77.00	162.51	1.55	620	! AA694773 GM01514.5prime GM Dros
gb_est17:AA637953	+	75.00	161.39	1.79	448	! AA637953 vr22c12.r1 Barstead mc
em_est21:HSM006748	-	75.00	161.21	1.83	455	! Al041898 Homo sapiens mRNA; EST
gb_est25:AI301307	-	74.00	160.67	1.96	386	! AI301307 qn45c07.x1 NCI_CGAP_Ki
gb_est26:AI378874	-	74.00	159.68	2.22	421	! AI378874 tc55b07.x1 Soares_NhHM
gb_est19:AA784252	-	73.00	157.86	2.81	400	! AA784252 d3f06a1.r1 Aspergillus
gb_est22:AI068978	-	76.00	157.09	3.10	807	! AI068978 mgae0004dH05f Magnapor
gb_est19:AA786837	-	72.00	155.45	3.83	400	! AA786837 m6e04a1.r1 Aspergillus
gb_est20:AA824336	+	74.00	154.88	4.12	642	! AA824336 aj30e05.s1 Soares_test
gb_est22:AI061719	-	74.00	154.70	4.21	652	! AI061719 LD34804.5prime LD Dros
gb_est16:AA540147	-	73.00	154.08	4.56	557	! AA540147 LD19533.5prime LD Dros
gb_est14:AA390962	-	73.00	153.98	4.62	562	! AA390962 LD09707.5prime LD Dros
gb_est11:AA263573	-	73.00	153.88	4.68	567	! AA263573 LD06751.5prime LD Dros
gb_est11:AA264527	-	73.00	153.86	4.69	568	! AA264527 LD07654.5prime LD Dros
gb_est2:T93447	+	70.00	153.27	5.05	317	! T93447 lmEST0217 LmLV39cDNA Lei
gb_est19:AA264805	-	73.00	152.94	5.28	616	! AA264805 LD08603.5prime LD D
gb_est19:AA390478	-	73.00	152.92	5.29	617	! AA390478 LD08741.5prime LD D
gb_est23:AI133933	-	72.50	152.32	5.71	585	! AI133933 GH11211.5prime GH D

seq_name: gb_est9:AA107012

seq_documentation_block:
LOCUS AA107012 377 bp mRNA EST 04-FEB-1997
DEFINITION ml93g04.r1 Stratagene mouse kidney (#937315) Mus musculus cDNA clone IMAGE:519606 5' similar to TR:E204404 E204404 TGR-CL1BIS ;, mRNA sequence.
ACCESSION AA107012
NID g1657080
VERSION AA107012.1 GI:1657080
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 377)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1404860.

Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:313454
Seq primer: -28m13 rev1 ET from Amersham
High quality sequence stop: 335.

FEATURES Location/Qualifiers
source 1. .377
/organism="Mus musculus"
/strain="C57/B16"
/db_xref="taxon:10090"
/map="12q13"
/clone="IMAGE:519606"
/clone_lib="Stratagene mouse kidney (#937315)"
/sex="females"
/tissue_type="kidney"
/dev_stage="4 weeks"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: kidney; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dT. Average insert size: 1.0 kb; Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT 68 a 95 c 143 g 71 t
ORIGIN

alignment_scores:
Quality: 90.50 Length: 152
Ratio: 1.207 Gaps: 5
Percent Similarity: 49.342 Percent Identity: 20.395

alignment_block:
US-09-068-804-5 x AA107012 ..

Align seg 1/1 to: AA107012 from: 1 to: 377

55 AlaAlaThrGlyValSerPheIleGlnGlnAlaLeuAsnProIleMetGl 71

```
      19 GCTGGAGGCAGTGTCCCTGGGATCGAGAGG...ATGGGCCCTGGCATTGA 65

      71 uHisValLeuLysProLeuMetGluLeuIleGlyLysAlaIleThrLysA 88
         :::::::      :::   ||||||   :::|||:::::::::      :
      66 CCGCATTAGCGGTGCTGGCATGGAGCGCATGGGCCGAGGCTTAGGCCATG 115

      88 laLeuGluGlyLeuGlyValAspArgLysArgGlnArgTrpProAlaAla 104
         ::::::::   :::|||   :::       |||
     116 GCATGGATCGAGTGGGCTCTGAGATTGAGCGC..................... 147

     105 LeuLeuValArgLeuSerProLeuCysHisGlyAspAlaValIleValVa 121

     147 ......................................................... 147

     121 lValAlaValValGlyLysGlyAlaAlaAlaLysLeuGlyAsnAlaLeuS 138
                                          :::|||   :::::::
     148 ..............................ATGGGCCTGGTCATGG 163

     138 erLysMet..............MetGlyGluThrIleLysLysLeuVal 149
         :::::|||               ||||||   :::|||   ::::::
     164 ACCGCATGGGCTCAGTTGAGCGCATGGGCTCCAGCATTGAGCGCATGGGC 213

     150 ProAsnValLeuLysGlnLeuAlaGlnAsnGlySerLysLeuPheThrGl 166
         |||      |||   :::::::|||   :::      ::::::      ||
     214 CCACTAGGCCTTGACCACATGGCCTCCAGTATTGAGCGCATG...GGCCA 260

     166 nGlyMetGlnArgIleThrSerGlyLeuGlyAsnValGlySerLysMet. 182
         |   |||:::|||||||   ||||||:::      :::|||:::   |||
     261 GACCATGGAGCGCATTGGTTCTGGCGTGGAGCGCATGGGTGCCGGCATGG 310

     183 .....GlyLeuGlnThrAsnAlaLeuSerLysGluLeuValGlyAsnThr 197
              ||||||:::      |||   :::   :::   |||||:::::::
     311 GCTTCGGCCTGGAGCGCATGGCCGCACCCATTGACCGTGTGGGCAAACCA 360

     198 LeuAsn 199
         |||:::
     361 TTGAGC 366
```

seq_name: gb_est17:AA615768

seq_documentation_block:

```
LOCUS       AA615768      478 bp    mRNA            EST       07-OCT-1997
DEFINITION  vo72f06.r1 Barstead mouse myotubes MPLRB5 Mus musculus cDNA clone
            IMAGE:1064675 5' similar to TR:E204306 E204306 TGR-CL5BIS ;, mRNA
            sequence.
ACCESSION   AA615768
NID         g2502996
VERSION     AA615768.1  GI:2502996
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 478)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Sep 1, 1995 this sequence version replaced.

            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:587035
            Possible reversed clone: similarity on wrong strand
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 207.
FEATURES             Location/Qualifiers
     source          1. .478
                     /organism="Mus musculus"
                     /strain="C3H"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1064675"
                     /clone_lib="Barstead mouse myotubes MPLRB5"
                     /cell_line="C2C12"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     [AATTCGGATCCTTG], digested with Not I and cloned into the
                     Not I and Eco RI sites of the modified pT7T3 vector.
                     Library constructed by Bob Barstead. The C2C12 cell line
                     (available from ATCC, catalog # CRL-1772) differentiates
                     rapidly, forming contractile myotubes and producing
                     characteristic muscle proteins."
BASE COUNT       82 a    120 c    180 g     94 t      2 others
ORIGIN
```

```
alignment_scores:
            Quality:   88.00              Length:    141
              Ratio:  1.060                Gaps:      7
 Percent Similarity:  58.865   Percent Identity: 24.113

alignment_block:
 US-09-068-804-5 x AA615768   ..

 Align seg 1/1  to: AA615768  from: 1  to: 478

      89 LeuGluGlyLeuGlyValAspArgLysArgGlnArgTrpProAlaAlaLe 105
         ||||||:::::::::::::::   :::         |||:::   |||||
      17 CTGGAGGCAGTGTCCCTGGGATCGAGA......GGATGGGCCCTGGCATT 60

     105 u..................LeuValArgLeuSerProLeuCysHisGlyA 116
         |                  :::   |||:::                |||:
      61 GACCGCATTAGTGGCTGGCATGGAGCGCATG.............GGCG 95

     116 spAlaValIleValValValAlaValValGlyLysGlyAlaAlaAlaLys 132
         ::|||:::        ::::::   ||||||   ::::::   ::::::
      96 AGGCCTTAGGCCATGGCATGGATCGAGTGGGC...TCTGAGATTGAGCGC 142

     133 LeuGlyAsnAlaLeuSerLysMet..............MetGlyGluTh 144
         :::|||   ::::::::::::|||               ||||||   ::
     143 ATGGGCCTGGTCATGGACCGCATGGGCTCAGTTGAGCGCATGGGCTCCAG 192

     144 rIleLysLysLeuValProAsnValLeuLysGlnLeuAlaGlnAsnGlyS 161
         :|||   ::::::   |||      :::   ::::::|||   :::
     193 CATCGAGCGCATGGGCCCACTAGGCGTTGACCACATGGCCTCCAGTATTG 242

     161 erLysLeuPheThrGlnGlyMetGlnArgIleThrSerGlyLeuGlyAsn 177
           ::::::      |||   |||:::||||||   ||||||:::
     243 AGCGCATG...GGCCAGACCATGGAGCGCATTGGTTCTGGCGTGGAGCGC 289

     178 ValGlySerLysMetGlyLeu......GlnThrAsnAlaLeuSerLysGl 192
         :::|||:::   ||||||:::      :::      |||   :::   ::
     290 ATGGGTGCCGGCATGGGCTTCGGCTNGGAGCGCATGGCCGCACCCATTGA 339

     192 uLeuValGlyAsnThrLeuAsnLysValAlaLeuGlyMetGluValThrA 209
         :   ||||||:::|||:::::::::::::::::   |||:::|||      :
     340 CCGTGTGGGCCAAACCATTGAGCGCATGGGCTCTGGTGTAGAGCGCATGG 389

     209 snThrAlaAlaGlnSerAlaGly 216
         :::::|||   :::      |||
     390 GTCCTGCCATTGAACGCATGGGC 412
```

```
seq_name: gb_est18:AA697508

seq_documentation_block:
LOCUS       AA697508      558 bp    mRNA            EST       27-NOV-1998
DEFINITION  HL02674.5prime HL Drosophila melanogaster head BlueScript
            Drosophila melanogaster cDNA clone HL02674 5prime, mRNA sequence.
ACCESSION   AA697508
NID         g2700437
VERSION     AA697508.1  GI:2700437
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 558)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     On Oct 30, 1997 this sequence version replaced gi:2160959.

            Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 26  row: G  column: 2
            High quality sequence stop: 394.
FEATURES             Location/Qualifiers
     source          1. .558
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="HL02674"
                     /clone_lib="HL Drosophila melanogaster head BlueScript"
                     /sex="male and female"
                     /dev_stage="adult"
                     /lab_host="SOLR"
                     /note="Organ: head--brain & sensory organ; Vector:
                     BlueScript SK; Site_1: EcoRI; Site_2: XhoI; Constructed
                     using Stratagene ZAP-cDNA Synthesis kit. Oligo dT-primed
                     and directionally cloned at EcoRI and XhoI in BlueScript
                     SK(+/-)"
BASE COUNT      120 a    179 c    155 g    103 t      1 others
ORIGIN


alignment_scores:
            Quality:   88.00            Length:    140
              Ratio:  1.173              Gaps:      6
 Percent Similarity:  53.571   Percent Identity: 26.429

alignment_block:
 US-09-068-804-5 x AA697508/rev   ..

 Align seg 1/1  to reverse of: AA697508  from: 1  to: 558

      96 ArgLysArgGlnArgTrpProAlaAla.LeuLeuValArgLeuSerProL 112
         |||:::      ||||||   :::||| |||||||||   |||
     468 CGCCGGCCAGGAAGGTGGCGTCCTGCAGCTTCTGGTTGAGCTG..... 426

     112 euCysHisGlyAspAlaValIleValValValAlaValValGlyLysGly 128

     426 ..................................................... 426

     129 AlaAlaAlaLysLeuGlyAsnAlaLeuSerLysMetMetGlyGluThrIl 145
                   |||   ::::::|||:::   ::::::|||   :::::
     425 ...........CTGCAGCACAGCCTCGGCCTCTTGCTTGGCAGTGCTGT 388

     145 eLysLysLeuValProAsnValLeuLysGlnLeuAlaGlnAsnGlySerL 162
         :                  :::|||:::|||   ||||||:::|||

     387 T.............CTTCTGCTGCGGCAGAATGCCCAGCAGGGGGAAG 353

     162 ysLeuPheThrGlnGlyMetGlnArgIleThrSerGlyLeuGlyAsnVal 178
               |||   |||   :::   ::::::   |||   |||:::
     352 ACCCAGGCACAGGAGGCAGGCACAATCTCGTTGTCGGCGAAGGAGATCCA 303

     179 GlySerLysMetGlyLeuGlnThrAsnAlaLeuSerLysGlu......Le 193
               :::|||||||:::   :::|||||||||::::::      ||
     302 CTGCTGGACCTGGGCCTGCACGAAGGGGCACTTTCGCGNGCGCAGCTGCT 253

     193 uValGlyAsnThrLeu...AsnLysValAlaLeuGlyMetGluValThrA 209
         ||||||||:::   :::   :::   ::::::|||:::   :::|||   :
     252 CGTTGGCCAGCAGGTAGGCGATGGCATTGGACTCGCT...CAAGTACTGT 206

     209 snThrAlaAlaGlnSerAlaGlyGlyValAlaGluGlyValPheIleLys 225
         :::::|||:::      ||||||   ::::::   |||   |||::::::
     205 CTTCGGCGGTCTCAAAGGCGGGCACCTTGCCACCGGGGAACTTCTTGAGG 156

     226 AsnAlaSerGluAlaLeu 231
         |||::::::::::   |||
     155 AACTCAGCCGACTTGTTG 138

seq_name: gb_est18:AA697063

seq_documentation_block:
LOCUS       AA697063      626 bp    mRNA            EST       28-NOV-1998
DEFINITION  GM09027.5prime GM Drosophila melanogaster ovary BlueScript
            Drosophila melanogaster cDNA clone GM09027 5prime, mRNA sequence.
ACCESSION   AA697063
NID         g2699683
VERSION     AA697063.1  GI:2699683
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 626)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     On May 5, 1995 this sequence version replaced gi:798168.

            Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 90  row: C  column: 3
            High quality sequence stop: 530.
FEATURES             Location/Qualifiers
     source          1. .626
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /map="20"
                     /clone="GM09027"
                     /clone_lib="GM Drosophila melanogaster ovary BlueScript"
                     /sex="female"
                     /dev_stage="newly eclosed females: germarium-stage 6"
                     /lab_host="SOLR"
                     /note="Organ: ovary; Vector: BlueScript SK; Site_1: EcoRI;
                     Site_2: XhoI; Constructed using Stratagene ZAP-cDNA
                     Synthesis kit. Oligo dT-primed and directionally cloned at
                     EcoRI and XhoI in BlueScript SK(+/-)"
BASE COUNT      135 a    202 c    174 g    115 t
ORIGIN


alignment_scores:
            Quality:   86.00            Length:    140
              Ratio:  1.162              Gaps:      6
```

```
 Percent Similarity:  52.857   Percent Identity: 27.143

alignment_block:
 US-09-068-804-5 x AA697063/rev   ..

 Align seg 1/1  to reverse of: AA697063  from: 1  to: 626

      96 ArgLysArgGlnArgTrpProAlaAla.LeuLeuValArgLeuSerProL 112
         |||:::      ||||||   :::||| |||||||||   |||
     461 CGCCGGCCAGGAAGGTGGCGTCCTGCAGCTTCTGGTTGAGCTG..... 419

     112 euCysHisGlyAspAlaValIleValValValAlaValValGlyLysGly 128

     419 ................................................. 419

     129 AlaAlaAlaLysLeuGlyAsnAlaLeuSerLysMetMetGlyGluThrIl 145
                    |||   :::::::|||:::   :::::::|||   :::::
     418 ............CTGCAGCACAGCCTCGGCCTCTTGCTTGGCAGTGCTGT 381

     145 eLysLysLeuValProAsnValLeuLysGlnLeuAlaGlnAsnGlySerL 162
         :                  :::|||:::|||   ||||||:::|||
     380 T...............CTTCTGCTGCGGCAGGATGCCCAGCAGGGGGAAG 346

     162 ysLeuPheThrGlnGlyMetGlnArgIleThrSerGlyLeuGlyAsnVal 178
                 |||   |||   :::   ::::::   |||   |||:::
     345 ACCCAGGCACAGGAGGCAGGCACAATCTCGTTGTCGGCGAAGGAGATCCA 296

     179 GlySerLysMetGlyLeuGlnThrAsnAlaLeuSerLysGlu......Le 193
                 ||||||||:::   :::||||||||   :::      ||
     295 CTGCTGGACATGGGCCTGCACGAAGGGGCACTTTCCGACGCGCAGCTGCT 246

     193 uValGlyAsnThrLeu...AsnLysValAlaLeuGlyMetGluValThrA 209
         |||||||:::   :::   :::   :::::::|||:::   :::|||   :
     245 CGTTGGCCAGCAGGTAGGCGATGGCATTGGACTCGCT...CAAGTACTGT 199

     209 snThrAlaAlaGlnSerAlaGlyGlyValAlaGluGlyValPheIleLys 225
         :::::|||:::      ||||||   ::::::   |||   |||::::::
     198 CTTCGGCGGTCTCAAAGGCGGGCACCTTGCCACCGGGGAACTTCTTGAGG 149

     226 AsnAlaSerGluAlaLeu 231
         |||:::::::::   |||
     148 AACTCAGCCGACTTGTTG 131

seq_name: gb_est16:AA541225

seq_documentation_block:
LOCUS       AA541225      613 bp    mRNA            EST       28-NOV-1998
DEFINITION  LD21073.5prime LD Drosophila melanogaster embryo BlueScript
            Drosophila melanogaster cDNA clone LD21073 5prime, mRNA sequence.
ACCESSION   AA541225
NID         g2287659
VERSION     AA541225.1  GI:2287659
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 613)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1394390.

            Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 210  row: G  column: 1
            High quality sequence stop: 255.

FEATURES             Location/Qualifiers
     source          1. .613
                     /organism="Drosophila melanogaster"
                     /db_xref="BDGP_EST:BDcln020100"
                     /db_xref="taxon:7227"
                     /clone="LD21073"
                     /clone_lib="LD Drosophila melanogaster embryo BlueScript"
                     /sex="male and female"
                     /dev_stage="0 to 24 hours mixed stage embryonic"
                     /lab_host="SOLR"
                     /note="Organ: embryo; Vector: BlueScript SK; Site_1:
                     EcoRI; Site_2: XhoI; Constructed using Stratagene ZAP-cDNA
                     Synthesis kit. Oligo dT-primed and directionally cloned at
                     EcoRI and XhoI in BlueScript SK(+/-)"
BASE COUNT      133 a    190 c    177 g    113 t
ORIGIN


alignment_scores:
           Quality:   85.00            Length:    187
             Ratio:  0.944              Gaps:      8
 Percent Similarity:  48.128   Percent Identity: 23.529

alignment_block:
 US-09-068-804-5 x AA541225/rev   ..

 Align seg 1/1  to reverse of: AA541225  from: 1  to: 613

      73 ValLeuLysProLeuMetGluLeuIle.....GlyLysAlaIl 85
         :::|||:::   |||   :::|||:::           ::::::
     602 ATCTTACGACGCCTGACCGCTCTGTTGAGATGTGACAACCAGCGTTCACG 553

      85 eThrLysAlaLeuGluGlyLeuGlyValAspArgLysArgGln..... 99
          :::   ||||||      |||:::   :::|||:::|||
     552 TTCCGAAGCACTGCGCACACTGGCTCCAGGACGTACTCGTACAGGTGGAG 503

      99 ................................................. 99

     502 CAGACTGCTGAAGACCACAATGTCGGCCAATGTGATCCTCTCGCCGGCCA 453

     100 ...ArgTrpProAlaAla.LeuLeuValArgLeuSerProLeuCysHisG 115
            ||||||   :::||| |||||||||   |||
     452 GGAAGGTGGCGTCCTGCAGCTTCTGGTTGAGCTG..............  419

     115 lyAspAlaValIleValValValAlaValValGlyLysGlyAlaAlaAla 131

     419 ................................................. 419

     132 LysLeuGlyAsnAlaLeuSerLysMetMetGlyGluThrIleLysLysLe 148
         |||   ::::::|||:::   ::::::|||   ::::::
     418 ...CTGCAGCACAGCCTCGGCCTCTTGCTTGGCAGTGCTGTT........ 380

     148 uValProAsnValLeuLysGlnLeuAlaGlnAsnGlySerLysLeuPheT 165
                   :::|||:::|||   ||||||:::|||           |
     379 .......CTTCTGCTGCGGCAGAATGCCCAGCAGGGGGAAGACCCAGGCA 337

     165 hrGlnGlyMetGlnArgIleThrSerGlyLeuGlyAsnValGlySerLys 181
         ||   |||   :::   ::::::   |||   |||:::
     336 CAGGAGGCAGGCACAATCTCGTTGTCGGCGAAGGAGATCCACTGCTGGAC 287

     182 MetGlyLeuGlnThrAsnAlaLeuSerLysGlu......LeuValGlyAs 196
         :::||||||:::   :::|||||||||   :::       |||||||||::
     286 CTGGGCCTGCACGAAGGGGCACTTTCCGCCGCGCAGCTGCTCGTTGGCCA 237

     196 nThrLeu...AsnLysValAlaLeuGlyMetGluValThrAsnThrAlaA 212
         :   :::   :::   ::::::|||:::   :::|||   ::::::|||:
     236 GCAGGTAGGCGATGGCATTGGACTCGCT...CAAGTACTGTCTTCGGCGG 190

     212 laGlnSerAlaGlyGlyValAlaGluGlyValPheIleLysAsnAlaSer 228
         ::      ||||||   ::::::   |||   |||::::::|||::::::
     189 TCTCAAAGGCGGGCACCTTGCCACCGGGGAACTTCTTGAGGAACTCAGCC 140

     229 GluAlaLeu 231
```

```
      :::   |||
  139 GACTTGTTG 131

seq_name: gb_est20:AA816702

seq_documentation_block:
LOCUS       AA816702      471 bp    mRNA            EST       25-NOV-1998
DEFINITION  LD04266.5prime LD Drosophila melanogaster embryo BlueScript
            Drosophila melanogaster cDNA clone LD04266 5prime, mRNA sequence.
ACCESSION   AA816702
NID         g2886310
VERSION     AA816702.1  GI:2886310
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 471)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 17, 1998 this sequence version replaced gi:2045525.

            Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 42  row: F  column: 6
            High quality sequence stop: 419.
FEATURES             Location/Qualifiers
     source          1. .471
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="LD04266"
                     /clone_lib="LD Drosophila melanogaster embryo BlueScript"
                     /sex="male and female"
                     /dev_stage="0 to 24 hours mixed stage embryonic"
                     /lab_host="SOLR"
                     /note="Organ: embryo; Vector: BlueScript SK; Site_1:
                     EcoRI; Site_2: XhoI; Constructed using Stratagene ZAP-cDNA
                     Synthesis kit. Oligo dT-primed and directionally cloned at
                     EcoRI and XhoI in BlueScript SK(+/-)"
BASE COUNT      104 a    153 c    135 g     79 t
ORIGIN


alignment_scores:
           Quality:   83.00            Length:    140
             Ratio:  1.122              Gaps:      6
 Percent Similarity:  52.857   Percent Identity: 26.429

alignment_block:
 US-09-068-804-5 x AA816702/rev   ..

 Align seg 1/1  to reverse of: AA816702  from: 1  to: 471

     96 ArgLysArgGlnArgTrpProAlaAla.LeuLeuValArgLeuSerProL 112
        |||:::     ||||||   :::||| ||||||||||   |||
    464 CGCCGGCCAGGAAGGTGGCGTCCTGCAGCTTCTGGTTGAGCTG..... 422

    112 euCysHisGlyAspAlaValIleValValValAlaValValGlyLysGly 128

    422 ..................................................... 422

    129 AlaAlaAlaLysLeuGlyAsnAlaLeuSerLysMetMetGlyGluThrIl 145
                |||   :::::::|||:::   :::::::|||   :::::
    421 ...........CTGCAGCACAGCCTCGGCCTCTTGCTTGGCAGTGCTGT 384

    145 eLysLysLeuValProAsnValLeuLysGlnLeuAlaGlnAsnGlySerL 162

        :                 :::|||:::|||   ||||||:::|||
    383 T..............CTTCTGCTGCGGCAGAATGCCCAGCAGGGGGAAG 349

    162 ysLeuPheThrGlnGlyMetGlnArgIleThrSerGlyLeuGlyAsnVal 178
            |||   |||   :::   ::::::   |||   |||:::
    348 ACCCAGGCACAGGAGGCAGGCACAATCTCGTTGTCGGCGAAGGAGATCCA 299

    179 GlySerLysMetGlyLeuGlnThrAsnAlaLeuSerLysGlu......Le 193
              :::||||||:::   :::|||||||||   :::      ||
    298 CTGCTGGACCTGGGCCTGCACGAAGGGGCACTTTCCGCCGCGCAGCTGCT 249

    193 uValGlyAsnThrLeu...AsnLysValAlaLeuGlyMetGluValThrA 209
        |||||||:::   :::   :::   ::::::|||:::   :::|||   :
    248 CGTTGGCCAGCAGGTAGGCGATGGCATTGGACTCGCT...CAAGTACTGT 202

    209 snThrAlaAlaGlnSerAlaGlyGlyValAlaGluGlyValPheIleLys 225
        :::::|||:::      ||||||   ::::::   |||   |||::::::
    201 CTTCGGCGGTCTCAAAGGCGGGCACCTTGCCACCGGGGAACTTCTTGAGG 152

    226 AsnAlaSerGluAlaLeu 231
        |||:::::::::   |||
    151 AACTCAGCCGACTTGTTG 134

seq_name: gb_est19:AA539284

seq_documentation_block:
LOCUS       AA539284      484 bp    mRNA            EST       27-NOV-1998
DEFINITION  LD18968.5prime LD Drosophila melanogaster embryo BlueScript
            Drosophila melanogaster cDNA clone LD18968 5prime, mRNA sequence.
ACCESSION   AA539284
NID         g2795051
VERSION     AA539284.1  GI:2795051
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 484)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2285729.

            Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 189  row: F  column: 8
            High quality sequence stop: 428.
FEATURES             Location/Qualifiers
     source          1. .484
                     /organism="Drosophila melanogaster"
                     /db_xref="BDGP_EST:BDcln018079"
                     /db_xref="taxon:7227"
                     /clone="LD18968"
                     /clone_lib="LD Drosophila melanogaster embryo BlueScript"
                     /sex="male and female"
                     /dev_stage="0 to 24 hours mixed stage embryonic"
                     /lab_host="SOLR"
                     /note="Organ: embryo; Vector: BlueScript SK; Site_1:
                     EcoRI; Site_2: XhoI; Constructed using Stratagene ZAP-cDNA
                     Synthesis kit. Oligo dT-primed and directionally cloned at
                     EcoRI and XhoI in BlueScript SK(+/-)"
BASE COUNT      110 a    156 c    136 g     82 t
ORIGIN


alignment_scores:
           Quality:   83.00            Length:    140
```

```
            Ratio:  1.122                Gaps:      6
Percent Similarity:  52.857   Percent Identity: 26.429

alignment_block:
 US-09-068-804-5 x AA539284/rev   ..

 Align seq 1/1  to reverse of: AA539284  from: 1  to: 484

      96 ArgLysArgGlnArgTrpProAlaAla.LeuLeuValArgLeuSerProL 112
         |||:::      ||||||   :::||| |||||||||   |||
     473 CGCCGGCCAGGAAGGTGGCGTCCTGCAGCTTCTGGTTGAGCTG..... 431

     112 euCysHisGlyAspAlaValIleValValValAlaValValGlyLysGly 128

     431 ..................................................... 431

     129 AlaAlaAlaLysLeuGlyAsnAlaLeuSerLysMetMetGlyGluThrIl 145
                 |||   ::::::|||:::   :::::::|||   :::::
     430 ...........CTGCAGCACAGCCTCGGCCTCTTGCTTGGCAGTGCTGT 393

     145 eLysLysLeuValProAsnValLeuLysGlnLeuAlaGlnAsnGlySerL 162
         :                 :::|||:::|||   |||||||:::|||
     392 T.............CTTCTGCTGCGGCAGAATGCCCAGCAGGGGGAAG 358

     162 ysLeuPheThrGlnGlyMetGlnArgIleThrSerGlyLeuGlyAsnVal 178
               |||   |||   :::   ::::::   |||   |||:::
     357 ACCCAGGCACAGGAGGCAGGCACAATCTCGTTGTCGGCGAAGGAGATCCA 308

     179 GlySerLysMetGlyLeuGlnThrAsnAlaLeuSerLysGlu......Le 193
               :::|||||:::   :::|||||||||   :::      ||
     307 CTGCTGGACCTGGGCCTGCACGAAGGGGCACTTTCCGCCGCGCAGCTGCT 258

     193 uValGlyAsnThrLeu...AsnLysValAlaLeuGlyMetGluValThrA 209
         |||||||:::   :::   :::   ::::::|||:::   :::|||   :
     257 CGTTGGCCAGCAGGTAGGCGATGGCATTGGACTCGCT...CAAGTACTGT 211

     209 snThrAlaAlaGlnSerAlaGlyGlyValAlaGluGlyValPheIleLys 225
         :::::|||:::      ||||||   ::::::   |||   |||::::::
     210 CTTCGGCGGTCTCAAAGGCGGGCACCTTGCCACCGGGGAACTTCTTGAGG 161

     226 AsnAlaSerGluAlaLeu 231
         |||:::::::::   |||
     160 AACTCAGCCGACTTGTTG 143

seq_name: gb_est11:AA264493

seq_documentation_block:
LOCUS       AA264493      636 bp    mRNA            EST       29-NOV-1998
DEFINITION  LD07632.5prime LD Drosophila melanogaster embryo BlueScript
            Drosophila melanogaster cDNA clone LD07632 5prime, mRNA sequence.
ACCESSION   AA264493
NID         g1900606
VERSION     AA264493.1  GI:1900606
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 636)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     On May 9, 1995 this sequence version replaced gi:802352.

            Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 76  row: C  column: 8

            High quality sequence stop: 559.
FEATURES             Location/Qualifiers
     source          1. .636
                     /organism="Drosophila melanogaster"
                     /db_xref="BDGP_EST:BDcln007003"
                     /db_xref="taxon:7227"
                     /clone="LD07632"
                     /clone_lib="LD Drosophila melanogaster embryo BlueScript"
                     /sex="male and female"
                     /dev_stage="0 to 24 hours mixed stage embryonic"
                     /lab_host="SOLR"
                     /note="Organ: embryo; Vector: BlueScript SK; Site_1:
                     EcoRI; Site_2: XhoI; Constructed using Stratagene ZAP-cDNA
                     Synthesis kit. Oligo dT-primed and directionally cloned at
                     EcoRI and XhoI in BlueScript SK(+/-)"
BASE COUNT      140 a    205 c    175 g    116 t
ORIGIN


alignment_scores:
            Quality:   84.00              Length:    140
              Ratio:  1.135                Gaps:      6
Percent Similarity:  52.857   Percent Identity: 26.429

alignment_block:
 US-09-068-804-5 x AA264493/rev   ..

 Align seq 1/1  to reverse of: AA264493  from: 1  to: 636

      96 ArgLysArgGlnArgTrpProAlaAla.LeuLeuValArgLeuSerProL 112
         |||:::      ||||||   :::||| |||||||||   |||
     472 CGCCGGCCAGGAAGGTGGCGTCCTGCAGCTTCTGGTTGAGCTG..... 430

     112 euCysHisGlyAspAlaValIleValValValAlaValValGlyLysGly 128

     430 ..................................................... 430

     129 AlaAlaAlaLysLeuGlyAsnAlaLeuSerLysMetMetGlyGluThrIl 145
                 |||   ::::::|||:::   :::::::|||   :::::
     429 ...........CTGCAGCACAGCCTCGGCCTCTTGCTTGGCAGTGCTGT 392

     145 eLysLysLeuValProAsnValLeuLysGlnLeuAlaGlnAsnGlySerL 162
         :                 :::|||:::|||   |||||||:::|||
     391 T.............CTTCTGCTGCGGCAGAATGCCCAGCAGGGGGAAG 357

     162 ysLeuPheThrGlnGlyMetGlnArgIleThrSerGlyLeuGlyAsnVal 178
               |||   |||   :::   ::::::   |||   |||:::
     356 ACCCAGGCACAGGAGGCAGGCACAATCTCGTTGTCGGCGAAGGAGATCCA 307

     179 GlySerLysMetGlyLeuGlnThrAsnAlaLeuSerLysGlu......Le 193
               :::|||||:::   :::|||||||||   :::      ||
     306 CTGCTGGACCTGGGCCTGCACGAAGGGGCACTTTCCAGCGCGCAGCTGCT 257

     193 uValGlyAsnThrLeu...AsnLysValAlaLeuGlyMetGluValThrA 209
         |||||||:::   :::   :::   ::::::|||:::   :::|||   :
     256 CGTTGGCCAGCAGGTAGGCGATGGCATTGGACTCGCT...CAAGTACTGT 210

     209 snThrAlaAlaGlnSerAlaGlyGlyValAlaGluGlyValPheIleLys 225
         :::::|||:::      ||||||   ::::::   |||   |||::::::
     209 CTTCGGCGGTCTCAAAGGCGGGCACCTTGCCACCGGGGAACTTCTTGAGG 160

     226 AsnAlaSerGluAlaLeu 231
         |||:::::::::   |||
     159 AACTCAGCCGACTTGTTG 142

seq_name: gb_est20:AA817581

seq_documentation_block:
LOCUS       AA817581      558 bp    mRNA            EST       25-NOV-1998
DEFINITION  LD23545.5prime LD Drosophila melanogaster embryo pOT2 Drosophila
            melanogaster cDNA clone LD23545 5prime, mRNA sequence.
ACCESSION   AA817581
NID         g2887190
```

```
VERSION     AA817581.1  GI:2887190
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 558)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2151218.

            Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 235  row: D  column: 9
            High quality sequence stop: 447.
FEATURES             Location/Qualifiers
     source          1. .558
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /map="21q"
                     /clone="LD23545"
                     /clone_lib="LD Drosophila melanogaster embryo pOT2"
                     /sex="male and female"
                     /dev_stage="0 to 24 hours mixed stage embryonic"
                     /lab_host="XL1 Blue"
                     /note="Organ: embryo; Vector: pOT2; Site_1: EcoRI; Site_2:
                     XhoI; Sized fractionated cDNAs were directly ligated into
                     pOT2. "
BASE COUNT      121 a    178 c    158 g    101 t
ORIGIN


alignment_scores:
            Quality:   83.00            Length:    140
              Ratio:  1.122              Gaps:      6
 Percent Similarity:  52.857   Percent Identity: 26.429

alignment_block:
 US-09-068-804-5 x AA817581/rev   ..

 Align seg 1/1  to reverse of: AA817581  from: 1  to: 558

      96 ArgLysArgGlnArgTrpProAlaAla.LeuLeuValArgLeuSerProL 112
         |||:::      ||||||   :::||| |||||||||   |||
     473 CGCCGGCCAGGAAGGTGGCGTCCTGCAGCTTCTGGTTGAGCTG..... 431

     112 euCysHisGlyAspAlaValIleValValValAlaValValGlyLysGly 128

     431 ..................................................... 431

     129 AlaAlaAlaLysLeuGlyAsnAlaLeuSerLysMetMetGlyGluThrIl 145
                 |||   ::::::|||:::   ::::::|||   :::::
     430 ...........CTGCAGCACAGCCTCGGCCTCTTGCTTGGCAGTGCTGT 393

     145 eLysLysLeuValProAsnValLeuLysGlnLeuAlaGlnAsnGlySerL 162
         :           :::|||:::|||   ||||||:::|||
     392 T.............CTTCTGCTGCGGCAGAATGCCCAGCAGGGGGAAG 358

     162 ysLeuPheThrGlnGlyMetGlnArgIleThrSerGlyLeuGlyAsnVal 178
               |||   |||   :::   ::::::   |||   |||:::
     357 ACCCAGGCACAGGAGGCAGGCACAATCTCGTTGTCGGCGAAGGAGATCCA 308

     179 GlySerLysMetGlyLeuGlnThrAsnAlaLeuSerLysGlu......Le 193
               :::|||||:::   :::|||||||||   :::      ||
     307 CTGCTGGACCTGGGCCTGCACGAAGGGGCACTTTCCGCCGCGCAGCTGCT 258

     193 uValGlyAsnThrLeu...AsnLysValAlaLeuGlyMetGluValThrA 209
         |||||||:::   :::   :::   ::::::|||:::   :::|||   :
     257 CGTTGGCCAGCAGGTAGGCGATGGCATTGGACTCGCT...CAAGTACTGT 211

     209 snThrAlaAlaGlnSerAlaGlyGlyValAlaGluGlyValPheIleLys 225
         :::::|||:::      ||||||   ::::::   |||   |||::::::
     210 CTTCGGCGGTCTCAAAGGCGGGCACCTTGCCACCGGGGAACTTCTTGAGG 161

     226 AsnAlaSerGluAlaLeu 231
         |||::::::::::   |||
     160 AACTCAGCCGACTTGTTG 143

seq_name: gb_est19:AA540277

seq_documentation_block:
LOCUS       AA540277      564 bp    mRNA            EST       27-NOV-1998
DEFINITION  LD19717.5prime LD Drosophila melanogaster embryo BlueScript
            Drosophila melanogaster cDNA clone LD19717 5prime, mRNA sequence.
ACCESSION   AA540277
NID         g2795309
VERSION     AA540277.1  GI:2795309
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 564)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2286711.

            Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 197  row: B  column: 5
            High quality sequence stop: 554.
FEATURES             Location/Qualifiers
     source          1. .564
                     /organism="Drosophila melanogaster"
                     /db_xref="BDGP_EST:BDcln018796"
                     /db_xref="taxon:7227"
                     /map="794G10; 7"
                     /clone="LD19717"
                     /clone_lib="LD Drosophila melanogaster embryo BlueScript"
                     /sex="male and female"
                     /dev_stage="0 to 24 hours mixed stage embryonic"
                     /lab_host="SOLR"
                     /note="Organ: embryo; Vector: BlueScript SK; Site_1:
                     EcoRI; Site_2: XhoI; Constructed using Stratagene ZAP-cDNA
                     Synthesis kit. Oligo dT-primed and directionally cloned at
                     EcoRI and XhoI in BlueScript SK(+/-)"
BASE COUNT      120 a    180 c    161 g    102 t      1 others
ORIGIN


alignment_scores:
            Quality:   83.00            Length:    140
              Ratio:  1.122              Gaps:      6
 Percent Similarity:  52.857   Percent Identity: 25.714

alignment_block:
 US-09-068-804-5 x AA540277/rev   ..

 Align seg 1/1  to reverse of: AA540277  from: 1  to: 564

      96 ArgLysArgGlnArgTrpProAlaAla.LeuLeuValArgLeuSerProL 112
         |||:::      ||||||   :::||| |||||||||   |||
```

```
     465 CGCCGGCCAGGAAGGTGGCGTCCTGCAGCTTCTGGTTGAGCTG..... 423

     112 euCysHisGlyAspAlaValIleValValValAlaValValGlyLysGly 128

     423 ................................................. 423

     129 AlaAlaAlaLysLeuGlyAsnAlaLeuSerLysMetMetGlyGluThrIl 145
                   |||   :::::::|||:::   :::::::|||   :::::
     422 ...........CTGCAGCACAGCCTCGGCCTCTTGCTTGGCAGTGCTGT 385

     145 eLysLysLeuValProAsnValLeuLysGlnLeuAlaGlnAsnGlySerL 162
         :                    :::|||:::|||   ||||||:::|||
     384 T..............CTTCTGCTGCGGCAGAATGCCCAGCAGGGGGAAG 350

     162 ysLeuPheThrGlnGlyMetGlnArgIleThrSerGlyLeuGlyAsnVal 178
               |||   |||   :::   ::::::   |||   |||:::
     349 ACCCAGGCACAGGAGGCAGGCACAATCTCGTTGTCGGCGAAGGAGATCCA 300

     179 GlySerLysMetGlyLeuGlnThrAsnAlaLeuSerLysGlu......Le 193
                 :::||||||:::   :::||||||:::   :::      ||
     299 CTGCTGGACCTGGGCCTGCACGAAGGGGCACTTCCCGACGCGCAGCTGCT 250

     193 uValGlyAsnThrLeu...AsnLysValAlaLeuGlyMetGluValThrA 209
         |||||||:::   :::   :::   :::::::|||:::   :::|||   :
     249 CGTTGGCCAGCAGGTAGGCGATGGCATTGGACTCGCT...CAAGTACTGT 203

     209 snThrAlaAlaGlnSerAlaGlyGlyValAlaGluGlyValPheIleLys 225
         :::::|||:::      ||||||   ::::::   |||   |||::::::
     202 CTTCGGCGGTCTCAAAGGCGGGCACCTTGCCACCGGGGAACTTCTTGAGG 153

     226 AsnAlaSerGluAlaLeu 231
         |||:::::::::   |||
     152 AACTCAGCCGACTTGTTG 135

seq_name: gb_est14:AA392365

seq_documentation_block:
LOCUS       AA392365      578 bp    mRNA            EST       28-NOV-1998
DEFINITION  LD11385.5prime LD Drosophila melanogaster embryo BlueScript
            Drosophila melanogaster cDNA clone LD11385 5prime, mRNA sequence.
ACCESSION   AA392365
NID         g2045493
VERSION     AA392365.1  GI:2045493
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 578)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1404998.

            Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 113  row: H  column: 1
            High quality sequence stop: 426.
FEATURES             Location/Qualifiers
     source          1. .578
                     /organism="Drosophila melanogaster"
                     /db_xref="BDGP_EST:BDcln010608"
                     /db_xref="taxon:7227"
                     /clone="LD11385"
                     /clone_lib="LD Drosophila melanogaster embryo BlueScript"
                     /sex="male and female"
                     /dev_stage="0 to 24 hours mixed stage embryonic"
                     /lab_host="SOLR"
                     /note="Organ: embryo; Vector: BlueScript SK; Site_1:
                     EcoRI; Site_2: XhoI; Constructed using Stratagene ZAP-cDNA
                     Synthesis kit. Oligo dT-primed and directionally cloned at
                     EcoRI and XhoI in BlueScript SK(+/-)"
BASE COUNT      122 a    184 c    161 g    111 t
ORIGIN


alignment_scores:
            Quality:   83.00            Length:    140
              Ratio:  1.122              Gaps:      6
 Percent Similarity:  52.857   Percent Identity: 26.429

alignment_block:
 US-09-068-804-5 x AA392365/rev   ..

 Align seg 1/1  to reverse of: AA392365  from: 1  to: 578

      96 ArgLysArgGlnArgTrpProAlaAla.LeuLeuValArgLeuSerProL 112
         |||:::      ||||||   :::||| |||||||||   |||
     495 CGCCGGCCAGGAAGGTGGCGTCCTGCAGCTTCTGGTTGAGCTG..... 453

     112 euCysHisGlyAspAlaValIleValValValAlaValValGlyLysGly 128

     453 ................................................. 453

     129 AlaAlaAlaLysLeuGlyAsnAlaLeuSerLysMetMetGlyGluThrIl 145
                   |||   :::::::|||:::   :::::::|||   :::::
     452 ...........CTGCAGCACAGCCTCGGCCTCTTGCTTGGCAGTGCTGT 415

     145 eLysLysLeuValProAsnValLeuLysGlnLeuAlaGlnAsnGlySerL 162
         :                    :::|||:::|||   ||||||:::|||
     414 T..............CTTCTGCTGCGGCAGAATGCCCAGCAGGGGGAAG 380

     162 ysLeuPheThrGlnGlyMetGlnArgIleThrSerGlyLeuGlyAsnVal 178
               |||   |||   :::   ::::::   |||   |||:::
     379 ACCCAGGCACAGGAGGCAGGCACAATCTCGTTGTCGGCGAAGGAGATCCA 330

     179 GlySerLysMetGlyLeuGlnThrAsnAlaLeuSerLysGlu......Le 193
                 :::||||||:::   :::||||||||||   :::      ||
     329 CTGCTGGACCTGGGCCTGCACGAAGGGGCACTTTCCGCCGCGCAGCTGCT 280

     193 uValGlyAsnThrLeu...AsnLysValAlaLeuGlyMetGluValThrA 209
         |||||||:::   :::   :::   :::::::|||:::   :::|||   :
     279 CGTTGGCCAGCAGGTAGGCGATGGCATTGGACTCGCT...CAAGTACTGT 233

     209 snThrAlaAlaGlnSerAlaGlyGlyValAlaGluGlyValPheIleLys 225
         :::::|||:::      ||||||   ::::::   |||   |||::::::
     232 CTTCGGCGGTCTCAAAGGCGGGCACCTTGCCACCGGGGAACTTCTTGAGG 183

     226 AsnAlaSerGluAlaLeu 231
         |||:::::::::   |||
     182 AACTCAGCCGACTTGTTG 165

seq_name: gb_est19:AA438684

seq_documentation_block:
LOCUS       AA438684      632 bp    mRNA            EST       27-NOV-1998
DEFINITION  LD13011.5prime LD Drosophila melanogaster embryo BlueScript
            Drosophila melanogaster cDNA clone LD13011 5prime, mRNA sequence.
ACCESSION   AA438684
NID         g2793002
VERSION     AA438684.1  GI:2793002
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 632)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
```

```
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2150565.

            Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 130  row: A  column: 11
            High quality sequence stop: 631.
FEATURES             Location/Qualifiers
     source          1. .632
                     /organism="Drosophila melanogaster"
                     /db_xref="BDGP_EST:BDcln012166"
                     /db_xref="taxon:7227"
                     /map="21q"
                     /clone="LD13011"
                     /clone_lib="LD Drosophila melanogaster embryo BlueScript"
                     /sex="male and female"
                     /dev_stage="0 to 24 hours mixed stage embryonic"
                     /lab_host="SOLR"
                     /note="Organ: embryo; Vector: BlueScript SK; Site_1:
                     EcoRI; Site_2: XhoI; Constructed using Stratagene ZAP-cDNA
                     Synthesis kit. Oligo dT-primed and directionally cloned at
                     EcoRI and XhoI in BlueScript SK(+/-)"
BASE COUNT      138 a    203 c    176 g    115 t
ORIGIN


alignment_scores:
           Quality:   83.00            Length:    140
             Ratio:  1.122              Gaps:      6
 Percent Similarity:  52.857   Percent Identity: 26.429

alignment_block:
 US-09-068-804-5 x AA438684/rev   ..

 Align seg 1/1  to reverse of: AA438684  from: 1  to: 632

      96 ArgLysArgGlnArgTrpProAlaAla.LeuLeuValArgLeuSerProL 112
         |||:::     ||||||   :::||| |||||||||   |||
     467 CGCCGGCCAGGAAGGTGGCGTCCTGCAGCTTCTGGTTGAGCTG..... 425

     112 euCysHisGlyAspAlaValIleValValValAlaValValGlyLysGly 128

     425 ................................................. 425

     129 AlaAlaAlaLysLeuGlyAsnAlaLeuSerLysMetMetGlyGluThrIl 145
                  |||   :::::::|||:::   :::::::|||   :::::
     424 ...........CTGCAGCACAGCCTCGGCCTCTTGCTTGGCAGTGCTGT 387

     145 eLysLysLeuValProAsnValLeuLysGlnLeuAlaGlnAsnGlySerL 162
         :                ::::|||:::|||   ||||||:::|||
     386 T..............CTTCTGCTGCGGCAGAATGCCCAGCAGGGGGAAG 352

     162 ysLeuPheThrGlnGlyMetGlnArgIleThrSerGlyLeuGlyAsnVal 178
               |||   |||   :::   ::::::   |||   |||:::
     351 ACCCAGGCACAGGAGGCAGGCACAATCTCGTTGTCGGCGAAGGAGATCCA 302

     179 GlySerLysMetGlyLeuGlnThrAsnAlaLeuSerLysGlu......Le 193
               :::|||||||:::   :::|||||||||   :::     ||
     301 CTGCTGGACCTGGGCCTGCACGAAGGGGCACTTTCCGCCGCGCAGCTGCT 252

     193 uValGlyAsnThrLeu...AsnLysValAlaLeuGlyMetGluValThrA 209
         |||||||:::   :::   :::   ::::::|||:::   :::|||   :
     251 CGTTGGCCAGCAGGTAGGCGATGGCATTGGACTCGCT...CAAGTACTGT 205

     209 snThrAlaAlaGlnSerAlaGlyGlyValAlaGluGlyValPheIleLys 225
         :::::|||:::      ||||||   ::::::   |||   |||::::::
     204 CTTCGGCGGTCTCAAAGGCGGGCACCTTGCCACCGGGGAACTTCTTGAGG 155

      226 AsnAlaSerGluAlaLeu 231
          |||::::::::::   |||
      154 AACTCAGCCGACTTGTTG 137

seq_name: gb_est21:AA978975

seq_documentation_block:
LOCUS       AA978975      698 bp    mRNA            EST       24-NOV-1998
DEFINITION  LD33203.5prime LD Drosophila melanogaster embryo pOT2 Drosophila
            melanogaster cDNA clone LD33203 5prime, mRNA sequence.
ACCESSION   AA978975
NID         g3156380
VERSION     AA978975.1  GI:3156380
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 698)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 17, 1998 this sequence version replaced gi:2044705.

            Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 332  row: A  column: 3
            High quality sequence stop: 596.
FEATURES             Location/Qualifiers
     source          1. .698
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /map="13"
                     /clone="LD33203"
                     /clone_lib="LD Drosophila melanogaster embryo pOT2"
                     /sex="male and female"
                     /dev_stage="0 to 24 hours mixed stage embryonic"
                     /lab_host="XL1 Blue"
                     /note="Organ: embryo; Vector: pOT2; Site_1: EcoRI; Site_2:
                     XhoI; Sized fractionated cDNAs were directly ligated into
                     pOT2. "
BASE COUNT      155 a    223 c    197 g    123 t
ORIGIN


alignment_scores:
           Quality:   83.00            Length:    140
             Ratio:  1.122              Gaps:      6
 Percent Similarity:  52.857   Percent Identity: 26.429

alignment_block:
 US-09-068-804-5 x AA978975/rev   ..

 Align seg 1/1  to reverse of: AA978975  from: 1  to: 698

      96 ArgLysArgGlnArgTrpProAlaAla.LeuLeuValArgLeuSerProL 112
         |||:::     ||||||   :::||| |||||||||   |||
     477 CGCCGGCCAGGAAGGTGGCGTCCTGCAGCTTCTGGTTGAGCTG..... 435

     112 euCysHisGlyAspAlaValIleValValValAlaValValGlyLysGly 128

     435 ................................................. 435

     129 AlaAlaAlaLysLeuGlyAsnAlaLeuSerLysMetMetGlyGluThrIl 145
                  |||   :::::::|||:::   :::::::|||   :::::
     434 ...........CTGCAGCACAGCCTCGGCCTCTTGCTTGGCAGTGCTGT 397
```

```
    145 eLysLysLeuValProAsnValLeuLysGlnLeuAlaGlnAsnGlySerL 162
        :                   :::|||:::|||   |||||||:::|||
    396 T..............CTTCTGCTGCGGCAGAATGCCCAGCAGGGGGAAG 362

    162 ysLeuPheThrGlnGlyMetGlnArgIleThrSerGlyLeuGlyAsnVal 178
              |||   |||   :::   ::::::   |||   |||:::
    361 ACCCAGGCACAGGAGGCAGGCACAATCTCGTTGTCGGCGAAGGAGATCCA 312

    179 GlySerLysMetGlyLeuGlnThrAsnAlaLeuSerLysGlu......Le 193
              :::|||||||:::   :::|||||||||   :::      ||
    311 CTGCTGGACCTGGGCCTGCACGAAGGGGCACTTTCCGCCGCGCAGCTGCT 262

    193 uValGlyAsnThrLeu...AsnLysValAlaLeuGlyMetGluValThrA 209
        |||||||:::   :::   :::   ::::::|||:::   :::|||   :
    261 CGTTGGCCAGCAGGTAGGCGATGGCATTGGACTCGCT...CAAGTACTGT 215

    209 snThrAlaAlaGlnSerAlaGlyGlyValAlaGluGlyValPheIleLys 225
        :::::|||:::      ||||||   ::::::   |||   |||::::::
    214 CTTCGGCGGTCTCAAAGGCGGGCACCTTGCCACCGGGGAACTTCTTGAGG 165

    226 AsnAlaSerGluAlaLeu 231
        |||:::::::::   |||
    164 AACTCAGCCGACTTGTTG 147
```

seq_name: gb_est21:AA979658

seq_documentation_block:

```
LOCUS       AA979658      740 bp    mRNA            EST       24-NOV-1998
DEFINITION  LD34364.5prime LD Drosophila melanogaster embryo pOT2 Drosophila
            melanogaster cDNA clone LD34364 5prime, mRNA sequence.
ACCESSION   AA979658
NID         g3157063
VERSION     AA979658.1  GI:3157063
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 740)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2150983.

            Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 343  row: F  column: 4
            High quality sequence stop: 679.
FEATURES             Location/Qualifiers
     source          1. .740
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /map="21q"
                     /clone="LD34364"
                     /clone_lib="LD Drosophila melanogaster embryo pOT2"
                     /sex="male and female"
                     /dev_stage="0 to 24 hours mixed stage embryonic"
                     /lab_host="XL1 Blue"
                     /note="Organ: embryo; Vector: pOT2; Site_1: EcoRI; Site_2:
                     XhoI; Sized fractionated cDNAs were directly ligated into
                     pOT2. "
BASE COUNT      169 a    237 c    210 g    124 t
ORIGIN
```

alignment_scores:

```
            Quality:   83.00            Length:    140
            Ratio:  1.122               Gaps:      6
 Percent Similarity:  52.857   Percent Identity: 26.429
```

alignment_block:

```
 US-09-068-804-5 x AA979658/rev   ..

 Align seg 1/1  to reverse of: AA979658  from: 1  to: 740

      96 ArgLysArgGlnArgTrpProAlaAla.LeuLeuValArgLeuSerProL 112
         |||:::      ||||||   :::||| ||||||||   |||
     473 CGCCGGCCAGGAAGGTGGCGTCCTGCAGCTTCTGGTTGAGCTG..... 431

     112 euCysHisGlyAspAlaValIleValValValAlaValValGlyLysGly 128

     431 ..................................................... 431

     129 AlaAlaAlaLysLeuGlyAsnAlaLeuSerLysMetMetGlyGluThrIl 145
                    |||   ::::::|||:::   ::::::|||   :::::
     430 ............CTGCAGCACAGCCTCGGCCTCTTGCTTGGCAGTGCTGT 393

     145 eLysLysLeuValProAsnValLeuLysGlnLeuAlaGlnAsnGlySerL 162
         :                   :::|||:::|||   |||||||:::|||
     392 T..............CTTCTGCTGCGGCAGAATGCCCAGCAGGGGGAAG 358

     162 ysLeuPheThrGlnGlyMetGlnArgIleThrSerGlyLeuGlyAsnVal 178
               |||   |||   :::   ::::::   |||   |||:::
     357 ACCCAGGCACAGGAGGCAGGCACAATCTCGTTGTCGGCGAAGGAGATCCA 308

     179 GlySerLysMetGlyLeuGlnThrAsnAlaLeuSerLysGlu......Le 193
               :::|||||||:::   :::|||||||||   :::      ||
     307 CTGCTGGACCTGGGCCTGCACGAAGGGGCACTTTCCGCCGCGCAGCTGCT 258

     193 uValGlyAsnThrLeu...AsnLysValAlaLeuGlyMetGluValThrA 209
         |||||||:::   :::   :::   :::::::|||:::   :::|||   :
     257 CGTTGGCCAGCAGGTAGGCGATGGCATTGGACTCGCT...CAAGTACTGT 211

     209 snThrAlaAlaGlnSerAlaGlyGlyValAlaGluGlyValPheIleLys 225
         :::::|||:::      ||||||   ::::::   |||   |||::::::
     210 CTTCGGCGGTCTCAAAGGCGGGCACCTTGCCACCGGGGAACTTCTTGAGG 161

     226 AsnAlaSerGluAlaLeu 231
         |||:::::::::   |||
     160 AACTCAGCCGACTTGTTG 143
```

seq_name: gb_est21:AA979513

seq_documentation_block:

```
LOCUS       AA979513      753 bp    mRNA            EST       24-NOV-1998
DEFINITION  LD34126.5prime LD Drosophila melanogaster embryo pOT2 Drosophila
            melanogaster cDNA clone LD34126 5prime, mRNA sequence.
ACCESSION   AA979513
NID         g3156918
VERSION     AA979513.1  GI:3156918
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 753)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2150650.

            Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 341  row: C  column: 2
```

```
          High quality sequence stop: 633.
FEATURES             Location/Qualifiers
     source          1. .753
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /map="5 q13; 754A01; 12; 12q12-12q13.13; 21q"
                     /clone="LD34126"
                     /clone_lib="LD Drosophila melanogaster embryo pOT2"
                     /sex="male and female"
                     /dev_stage="0 to 24 hours mixed stage embryonic"
                     /lab_host="XL1 Blue"
                     /note="Organ: embryo; Vector: pOT2; Site_1: EcoRI; Site_2:
                     XhoI; Sized fractionated cDNAs were directly ligated into
                     pOT2. "
BASE COUNT      176 a    238 c    216 g    123 t
ORIGIN


alignment_scores:
           Quality:   83.00              Length:    140
             Ratio:  1.122                Gaps:      6
 Percent Similarity:  52.857   Percent Identity: 26.429

alignment_block:
 US-09-068-804-5 x AA979513/rev   ..

 Align seg 1/1  to reverse of: AA979513  from: 1  to: 753

     96 ArgLysArgGlnArgTrpProAlaAla.LeuLeuValArgLeuSerProL 112
        |||:::      ||||||   :::||| ||||||||||   |||
    467 CGCCGGCCAGGAAGGTGGCGTCCTGCAGCTTCTGGTTGAGCTG..... 425

    112 euCysHisGlyAspAlaValIleValValValAlaValValGlyLysGly 128

    425 ................................................. 425

    129 AlaAlaAlaLysLeuGlyAsnAlaLeuSerLysMetMetGlyGluThrIl 145
                ||| :::::::|||:::  ::::::|||  :::::
    424 ...........CTGCAGCACAGCCTCGGCCTCTTGCTTGGCAGTGCTGT 387

    145 eLysLysLeuValProAsnValLeuLysGlnLeuAlaGlnAsnGlySerL 162
        :                 :::|||:::|||  ||||||:::|||
    386 T.............CTTCTGCTGCGGCAGAATGCCCAGCAGGGGGAAG 352

    162 ysLeuPheThrGlnGlyMetGlnArgIleThrSerGlyLeuGlyAsnVal 178
              |||  |||  :::  ::::::  |||  |||:::
    351 ACCCAGGCACAGGAGGCAGGCACAATCTCGTTGTCGGCGAAGGAGATCCA 302

    179 GlySerLysMetGlyLeuGlnThrAsnAlaLeuSerLysGlu......Le 193
            :::|||||:::  :::|||||||||  :::     ||
    301 CTGCTGGACCTGGGCCTGCACGAAGGGGCACTTTCCGCCGCGCAGCTGCT 252

    193 uValGlyAsnThrLeu...AsnLysValAlaLeuGlyMetGluValThrA 209
        |||||||:::  :::  :::  ::::::|||:::  :::|||  :
    251 CGTTGGCCAGCAGGTAGGCGATGGCATTGGACTCGCT...CAAGTACTGT 205

    209 snThrAlaAlaGlnSerAlaGlyGlyValAlaGluGlyValPheIleLys 225
        :::::|||:::     ||||||  ::::::  |||  |||::::::
    204 CTTCGGCGGTCTCAAAGGCGGGCACCTTGCCACCGGGGAACTTCTTGAGG 155

    226 AsnAlaSerGluAlaLeu 231
        |||:::::::::  |||
    154 AACTCAGCCGACTTGTTG 137
```

This Page Blank (uspto)

OM of: US-09-068-804-5 to: Issued_Patents_NA:* out_format : pfs

Date: Oct 5, 1999 7:09 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09068804/runat_05101999_145135_10326/app_query.fasta.1
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=pam150
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=escore
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09068804
-NCPU=6 -ICPU=3 -WAIT -THREADS=1

Search information block:
Query: US-09-068-804-5
Query length: 290
Database: Issued_Patents_NA:*
Database sequences: 192659
Database length: 52021692
Search time (sec): 101.650000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ..
/cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:PCT-US91-09422-20	-	94.50	185.39	0.0062	242	
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-247-901C-1	+	108.50	183.37	0.0080	50341	!
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-541-033A-25	-	91.00	182.84	0.0085	1473	!
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-541-033A-23	-	91.00	182.58	0.0088	1506	!
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-541-033A-7	-	91.00	179.42	0.0132	1969	!
/cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:PCT-US91-09422-18	-	94.50	179.22	0.0136	409	
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-325-267A-1	-	95.00	179.02	0.0139	4614	!
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-541-033A-19	-	91.00	178.69	0.0146	2096	!
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-541-033A-3	-	91.00	178.67	0.0146	2099	!
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-541-033A-18	-	91.00	178.46	0.0150	2137	!
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-541-033A-1	-	91.00	178.44	0.0150	2140	!
/cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:PCT-US93-10500-1	+	92.00	174.00	0.0265	3828	
/cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:PCT-US91-01327-12	-	86.50	171.36	0.0372	155	
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-471-033-39	-	82.50	164.41	0.0909	1241	!
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-471-033-42	-	82.50	164.41	0.0909	1241	!
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-471-044-39	-	82.50	164.41	0.0909	1241	!
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-471-044-42	-	82.50	164.41	0.0909	1241	!
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-463-483A-39	-	82.50	164.41	0.0909	1241	!
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-463-483A-42	-	82.50	164.41	0.0909	1241	!
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-471-046A-39	-	82.50	164.41	0.0909	1241	!
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-471-046A-42	-	82.50	164.41	0.0909	1241	!
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-470-566B-39	-	82.50	164.41	0.0909	1241	!
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-470-566B-42	-	82.50	164.41	0.0909	1241	!
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-951-715A-2	-	87.50	164.34	0.0917	3468	!
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-459-448A-2	-	87.50	164.34	0.0917	3468	!
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-484-358-6	+	81.50	164.27	0.0924	1023	!
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-471-033-45	-	82.50	163.35	0.1041	1358	!
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-471-044-45	-	82.50	163.35	0.1041	1358	!
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-463-483A-45	-	82.50	163.35	0.1041	1358	!
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-471-046A-45	-	82.50	163.35	0.1041	1358	!
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-470-566B-45	-	82.50	163.35	0.1041	1358	!
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-971-096-3	-	81.50	163.18	0.1064	1123	!
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-175-096-3	-	81.50	163.18	0.1064	1123	!
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-471-033-27	-	82.50	163.08	0.1077	1389	!
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-471-044-27	-	82.50	163.08	0.1077	1389	!
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-463-483A-27	-	82.50	163.08	0.1077	1389	!
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-471-046A-27	-	82.50	163.08	0.1077	1389	!
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-470-566B-27	-	82.50	163.08	0.1077	1389	!
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-471-033-24	-	82.50	163.00	0.1089	1399	!
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-471-044-24	-	82.50	163.00	0.1089	1399	!
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-463-483A-24	-	82.50	163.00	0.1089	1399	!
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-471-046A-24	-	82.50	163.00	0.1089	1399	!
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-470-566B-24	-	82.50	163.00	0.1089	1399	
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-951-715A-4	-	86.50	161.93	0.1248	3468	
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-459-448A-4	-	86.50	161.93	0.1248	3468	

seq_name: /cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:PCT-US91-09422-20

seq_documentation_block:
; Sequence 20, Application PC/TUS9109422
; GENERAL INFORMATION:
; APPLICANT: Mulvihill, Eileen R.
; APPLICANT: Hagen, Frederick S.
; APPLICANT: Houamed, Khaled M.
; APPLICANT: Almers, Wolfhard
; TITLE OF INVENTION: G PROTEIN-COUPLED GLUTAMATE RECEPTORS
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend
; STREET: One Market Plaza, Steuart Street Tower
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105-1492
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/09422
; FILING DATE: 19911212
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/672,007
; FILING DATE: 18-MAR-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/648,481
; FILING DATE: 30-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/626,806
; FILING DATE: 12-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Parmelee, Steven W.
; REGISTRATION NUMBER: 31,990
; REFERENCE/DOCKET NUMBER: 13952-6PC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 467-9600
; TELEFAX: (415) 543-5043
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2426 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; IMMEDIATE SOURCE:
; CLONE: SR13
PCT-US91-09422-20

alignment_scores:
Quality: 94.50 Length: 251
Ratio: 0.836 Gaps: 11
Percent Similarity: 45.020 Percent Identity: 23.506

alignment_block:
US-09-068-804-5 x PCT-US91-09422-20/rev ..

Align seg 1/1 to reverse of: PCT-US91-09422-20 from: 1 to: 2426

6 ThrAsnArgIleMetGlyCysIleGlyLysValLeuGlyAlaLeuLeuTh 22
1431 GCAGCTCGTGCGGTTCTCTGTGGGCCGCATGTCGTAGGGGCAGGTCTTAC 1382

```
     22 rIleValSerValValAlaAlaValPheThrGlyGlyAlaSerLeuAlaL 39
        |   |||:::   |||      |||::::::   ||||||:::||||||:
   1381 AGGTGTAGCG...GTCCACTTGGTACTGGTACCCGGTGCAGGGCTCGCAG 1335

     39 euAlaAlaValGlyLeuAlaValMetValAlaAspGluIleValLysAla 55
        :::::|||::::::
   1334 TGCCAGCAGCAAGC..................................... 1321

     56 AlaThrGlyValSerPheIleGlnGlnAlaLeuAsnProIleMetGluHi 72
                                                          ||
   1320 .....................................................CA 1319

     72 sValLeuLysProLeuMetGluLeuIleGlyLysAlaIleThrLysAlaL 89
        |:::|||   :::|||:::   |||   |||                   |
   1318 TGCCCTTCACAGTCTTCTTTCGCTCCCCGGG.................C 1287

     89 euGluGlyLeuGlyValAspArgLysArgGln.............ArgTrp 101
        ||:::|||      :::|||      ||||||               |||
   1286 TGGCAGGGCAGACTGCAGATGGAGCGCGGCAGCTGCTGGCCACTCCCTGG 1237

    102 ProAlaAlaLeuLeuValArgLeuSerProLeuCysHisGlyAspAlaVa 118
        |||                |||   ||||||:::
   1236 CCA.................CTGCATCCGCTCTATTCTGAG........ 1213

    118 lIleValValValAlaValValGlyLysGlyAlaAlaAlaLysLeuGlyA 135
               |||   :::|||      ::::::::::::      ||||||
   1212 .......GTGCAGGTGGTCTGTCCACGAGCCGATGACCTTGTACTCGGCC 1170

    135 snAlaLeuSerLysMetMet.............................. 141
          |||:::::::::::::::
   1169 GAGCCATTGCGCAGTTGGTACTGGTAGATGTCGTAGCGCCCCGGTGCGTC 1120

    142 ..............GlyGluThrIleLysLysLeu.............. 148
                     |||      :::   :::
   1119 TCCGTTCTCATTGAAGGTTACAGGGTTCCCCGCAATGCCTGAGAAGTTGA 1070

    149 .ValProAsnValLeuLysGlnLeuAla..................... 157
         ||||||:::||||||||||||||||:::
   1069 CGTTCCTGATGTACTTAAGCAGCTGGGTGCCATCCACGGGGTCCATGCGA 1020

    158 ..............GlnAsnGlySerLysLeuPheThrGlnGlyMetGln 169
                       :::|||::::::::::   :::::::|||:::|||
   1019 GGGCAGAGTCCTACGCGGCCGGGACACAGGTCACGGTGCATGGCGTGCAG 970

    170 ArgIleThrSerGlyLeuGlyAsnValGlySerLysMetGlyLeuGlnTh 186
        |||::::::   |||:::   :::|||:::      :::   |||:::
    969 CGCGTGGCCCATGGCGTACACAGCGTCAATCACGAACTGCACCTTCCCCT 920

    186 rAsnAlaLeuSerLysGluLeu........................... 193
              :::::::::   |||
    919 CCTGCTCATAGGCCGAGTCCTGCCCGATGCGCTCTCGGTTGGTGCACTTC 870

    194 .....ValGlyAsnThrLeuAsnLysValAla......LeuGlyMetGlu 206
         |||::::::   |||::::::||||||      |||:::||||||
    869 TTGATGTGGCTTCCCTTCTTGAGCGCGTGGCGGCTCAACTTGCAATGGAA 820

    207 Val 207
        |||
    819 GTT 817
```

```
seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-247-901C-1

seq_documentation_block:
; Sequence 1, Application US/08247901C
; Patent No. 5750384
;  GENERAL INFORMATION:
;    APPLICANT:  Jacobs et al
;    TITLE OF INVENTION:  L5 SHUTTLE PHASMIDS
;    NUMBER OF SEQUENCES:  1
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Amster, Rothstein & Ebenstein
;      STREET:  90 Park Avenue
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10016
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch 1.44 Mb storage diskette
;      COMPUTER:  IBM PC Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  Word Processor (ASCII)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/247,901C
;      FILING DATE:  May 23, 1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/057,531
;      FILING DATE:  April 29, 1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bogosian, Elizabeth A
;      REGISTRATION NUMBER:  39,911
;      REFERENCE/DOCKET NUMBER:  96700/273
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 697-5995
;      TELEFAX:  (212) 286-0854 or 286-0082
;      TELEX:  TWX 710-581-4766
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  50341
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:
;      DESCRIPTION:  L5 shuttle phasmid sequence
;    HYPOTHETICAL:  No
;    ANTI-SENSE:
;    FRAGMENT TYPE:
;    ORIGINAL SOURCE:
;      ORGANISM:  L5 mycobacteriophage
;      STRAIN:
;      INDIVIDUAL ISOLATE:
;      DEVELOPMENTAL STAGE:
;      HAPLOTYPE:
;      TISSUE TYPE:
;      CELL TYPE:
;      CELL LINE:
;      ORGANELLE:
;    IMMEDIATE SOURCE:
;    POSITION IN GENOME:
;      CHROMOSOME/SEGMENT:
;    FEATURE:
;      NAME/KEY:
;      LOCATION:
;      IDENTIFICATION METHOD:
;      OTHER INFORMATION:
;    PUBLICATION INFORMATION:  No. 5750384e
;      AUTHORS:
;      TITLE:
;      JOURNAL:
;      VOLUME:
;      PAGES:
;      DATE:
;      DOCUMENT NUMBER:
;      FILING DATE:
;      PUBLICATION DATE:
;      RELEVANT RESIDUES IN SEQ ID NO:
US-08-247-901C-1


alignment_scores:
            Quality:  108.50              Length:    251
              Ratio:  0.719                Gaps:     12
 Percent Similarity:  60.159   Percent Identity:  22.709

alignment_block:
 US-09-068-804-5 x US-08-247-901C-1    ..
```

```
Align seg 1/1  to: US-08-247-901C-1  from: 1  to: 50341

     13 IleGlyLysValLeuGlyAlaLeuLeuThrIleValSerValValAlaAl 29
        ||||||:::|||||||||||:::|||   |||   ::::::   ::::::::
  16998 ATCGGCAACGTCCTCGGGACGCTGGGCACACAGCTCGCTCCCATCGTCAC 17047

     29 aValPheThrGlyGlyAlaSerLeuAlaLeuAlaAlaValGlyLeuAlaV 46
        :::::::|||       :::::::::          |||       |||::::
  17048 GGCGCTCACG......CCGGCCTTCCAGACGCTGGCGAGCACGCTCGGCA 17091

     46 alMetValAlaAspGluIleValLysAlaAlaThrGlyValSerPheIle 62
        ::|||:::::::::::::::::
  17092 CGATGCTCACCGGAGCCCTC.............................. 17111

     63 GlnGlnAlaLeuAsnProIleMetGluHisValLeuLysProLeuMetGl 79
           ||||||||||:::|||||||:::    :::|||
  17112 ...CAAGCTCTGGGTCCGATCCTGACTCAGGTC............GCTAC 17146

     79 uLeuIleGlyLysAlaIleThrLysAlaLeuGluGlyLeu.......... 92
        ||||||||||:::::::::::::::::|||||:::::|||
  17147 GTTGATCGGCACGACGCTGAACACGGCGCTGCAGGCTCTCCAGCCGATGC 17196

     93 .....GlyValAspArgLysArgGlnArgTrpProAlaAlaLeuLeuVal 107
           ::::::   :::      |||:::   ::::::::::|||:::::::
  17197 TGCCGTCGCTCATGCAGAGCTTCCAGCAGATCTCCGACGTACTGGTGACC 17246

    108 ArgLeuSerProLeuCysHisGlyAspAlaValIleValValValAlaVa 124
           |||:::|||      |||      |||:::   ::::::::::::
  17247 AGTCTGGCCCCG......CACATCCCGGCGCTGGCGACGGCCCTCGGCCA 17290

    124 lValGlyLysGlyAlaAlaAlaLysLeuGlyAsnAlaLeuSerLysMetM 141
        |||:::   ||||||:::   :::|||:::   ::::::        :
  17291 GGTCGCA...GGCGCGGTGCTGCAGCTCGCTCCGACGATCATCTCGACGT 17337

    141 etGlyGluThrIleLysLysLeuValProAsnVal............... 152
        ::      ::::::   :::||||||||||:::|||
  17338 TGGTTCCGGCGTTCGTTCAGTTGGTCCCAAAGGTCGCTGAGCTAGTTCCG 17387

    153 ...LeuLysGlnLeuAlaGlnAsnGlySerLysLeu.............. 163
           :::   :::|||:::|||:::   :::::|||
  17388 ACCATCGTCAACCTGGTCCAGTCGTTCGCCAACCTGATGCCGGTGGTTCT 17437

    164 ....PheThrGlnGlyMetGlnArgIleThrSerGlyLeuGlyAsnValG 179
            :::::::|||::::::      :::::::::::::::   :::||||
  17438 GCCCCTGGCGCAGGCTCTGGTCAGCGTTGCTGGCGCGGTGATTCAGGTGG 17487

    179 lySerLysMet...GlyLeuGlnThrAsnAlaLeuSerLys......Glu 192
        ||      :::   |||       ::::::||||||::::::      |||
  17488 GTGTCTCCATCGGCGGCGCGCTCATCGGCGCGCTGGCGAACCTCACGGAG 17537

    193 LeuValGlyAsnThrLeuAsnLysVal....................... 201
        ::::::::::|||::::::::||||||
  17538 ATCATCTCCAACGTCATCAAGAAGGTGTCCGAGTGGGTCAGCAGCTTCTC 17587

    202 .AlaLeuGlyMetGluValThrAsnThrAlaAlaGlnSerAlaGlyGlyV 218
         :::   :::   ::::::::::::::::||||||:::   :::|||   :
  17588 CAGCGGAGCCCAGCAGATCGCTGCGAAGGCAGCGGAACTGCCGGGGATGA 17637

    218 alAlaGluGlyValPheIleLysAsnAlaSerGluAlaLeuAlaAspPhe 234
        ::                                |||||||||:::::
  17638 TC.........................CAGTCGGCTCTCGCCAACCTG 17660

    235 Met 235
        |||
  17661 ATG 17663

seq_name: /cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-541-033A-25

seq_documentation_block:
; Sequence 25, Application US/08541033A
; Patent No. 5879941
;  GENERAL INFORMATION:
;    APPLICANT:  Schmidt, Robert R.
;    APPLICANT:  Miller, Philip
;    TITLE OF INVENTION:  NOVEL POLYPEPTIDES AND POLYNUCLEOTIDES
;    TITLE OF INVENTION:  RELATING TO THE `- AND  -SUBUNITS OF GLUTAMATE
;    TITLE OF INVENTION:  DEHYDROGENASES AND METHODS OF USE
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Saliwanchik & Saliwanchik
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  Florida
;      COUNTRY:  USA
;      ZIP:  32606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/541,033A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Whitlock, Ted W.
;      REGISTRATION NUMBER:  36,965
;      REFERENCE/DOCKET NUMBER:  UF155
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (904) 375-8100
;      TELEFAX:  (904) 372-5800
;  INFORMATION FOR SEQ ID NO:  25:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1473 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  CDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  4..1431
US-08-541-033A-25


alignment_scores:
            Quality:   91.00              Length:    267
              Ratio:  0.615                Gaps:     15
 Percent Similarity:  55.431   Percent Identity: 21.348

alignment_block:
 US-09-068-804-5 x US-08-541-033A-25/rev   ..

 Align seg 1/1  to reverse of: US-08-541-033A-25  from: 1  to: 1473

      17 LeuGlyAlaLeuLeuThrIleValSerValValAlaAlaValPheThrGl 33
         |||::::::   |||:::   ::::::::::|||||||   :::      ||
    1035 CTGGCAGTCCAGCTCCCAAGGCTTGCGGCGGTCGCCCACATA...CACGG 989

      33 yGlyAlaSerLeuAlaLeuAlaAlaValGlyLeuAlaValMetValAlaA 50
         ||||:::::::|||:::|||::::::::::|||   :::|||::::::   :
     988 CGGTGTCGCTCTTGTACTCGGAGATGCGGGCGCTGTTGTTCTTCTTCTTC 939

      50 spGluIleValLys.............AlaAlaThrGlyValSerPheIle 62
         ::   :::   :::              |||       :::|||::::::::::
     938 ATGTCCTGCACCGCCTGCAGCTGCTCGCGCGTGAAGCCGTTGGGCTCGTA 889

      63 GlnGlnAlaLeu.......................AsnProIleMetGl 71
         :::   ||||||                       ::::::::::::::
     888 CACGTAGCCCTGGGAGTCGGACAGCGACAGCACGATGGCGCCCTTCTCCA 839

      71 uHisValLeuLysProLeuMetGluLeuIleGlyLysAla.........I 85
         ::::   |||::::::::::::      :::::::::|||
     838 GCAGCAGCTCCGCGCAGTACTGGGCCACGTTGCCCGCGCCAGACACCAGG 789

      85 leThrLysAlaLeuGluGlyLeuGlyVal..................... 94
```

```
                :::   ||||||||||:::|||::::::
    788 CAGCGCTTGCCCTTGAGGCTCTCGCCCTTGTCCTTCAGCACGTTCTCCAC 739

     95 ...Asp..ArgLysArgGlnArgTrpPro.AlaAlaLeuLeuValArgLe 109
           :::  |||:::|||   ||||||||| :::|||   |||
    738 AAACAGCACGGCGCCGTAGCCGGTGGCCTCGGGGCGGATCTC........ 697

    109 uSerProLeuCysHisGlyAspAlaValIleValValValAlaValValG 126
                              ::::::   :::::::::|||:::   |
    696 .....................GGAGCCGCCATACTCCTGGCCCTTCGGG 669

    126 lyLysGlyAlaAlaAlaLysLeuGlyAsnAlaLeuSerLysMetMetGly 142
        ||:::   |||:::::::   ||||||:::|||||||      :::::::::
    668 GTCAGCACGCCGGTGTAGTTCTTGGTGATGCGCTT......GTACTGGCC 625

    143 GluThrIleLysLysLeuValProAsnValLeuLysGlnLeuAlaGlnAs 159
        |||::::::   :::|||
    624 GAAAAGGTAGCCAATCTC.............................. 607

    159 nGlySerLysLeuPheThrGlnGlyMetGlnArgIleThrSerGlyLeuG 176
                                                  :::   :::   :
    606 ....................................GCGCGCGCCCACG 594

    176 lyAsnValGlySerLysMetGlyLeuGlnThrAsnAlaLeuSerLysGlu 192
        :::::|||::::::      |||:::::::::::::::::::   :::
    593 CCGATGTCGCCGGCGGGCACGTCCTGCACGTAGCTGATGTGGCGCTGCAG 544

    193 LeuValGlyAsnThrLeuAsn..............LysValAlaLeuGl 204
        |||      :::   |||:::                ::::::::::::
    543 CTCGGTCATGAAGGACTGGCAGAAGCGCATCACCTCCGCGTCGCTCTTGC 494

    204 yMetGluValThrAsnThrAlaAlaGlnSerAlaGly...GlyValAlaG 220
        ::::   |||         ||||||   :::|||:::   |||   :::
    493 CCTTGGGGTCGAAGTCGGAGCCGCCCTTGCCGCCGCCCATGGGCAGGGTG 444

    220 luGly.ValPheIleLysAsnAlaSerGluAlaLeuAlaAspPheMetLe 236
           :::|||:::|||      |||   |||      :::||||||::
    443 GTCAGGCTGTTCTTGAAGATCTGCTCAAAGGCA...AGGAACTTCATGAT 397

    236 uAlaArgPheAlaMetAspGlnIleGlnGlnTrpLeuLysGlnSer 251
        ::::|||||||:::            :::   |||   :::::::::
    396 GGACAGGTTCACG............GAGGGGTGGAAGCGCAGGCCG 363

seq_name: /cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-541-033A-23

seq_documentation_block:
; Sequence 23, Application US/08541033A
; Patent No. 5879941
;  GENERAL INFORMATION:
;    APPLICANT:  Schmidt, Robert R.
;    APPLICANT:  Miller, Philip
;    TITLE OF INVENTION:  NOVEL POLYPEPTIDES AND POLYNUCLEOTIDES
;    TITLE OF INVENTION:  RELATING TO THE '- AND  -SUBUNITS OF GLUTAMATE
;    TITLE OF INVENTION:  DEHYDROGENASES AND METHODS OF USE
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Saliwanchik & Saliwanchik
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  Florida
;      COUNTRY:  USA
;      ZIP:  32606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/541,033A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Whitlock, Ted W.
;      REGISTRATION NUMBER:  36,965
;      REFERENCE/DOCKET NUMBER:  UF155
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (904) 375-8100
;      TELEFAX:  (904) 372-5800
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1506 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  CDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  4..1464
US-08-541-033A-23


alignment_scores:
            Quality:   91.00              Length:    267
              Ratio:  0.615                Gaps:     15
 Percent Similarity:  55.431   Percent Identity: 21.348

alignment_block:
 US-09-068-804-5 x US-08-541-033A-23/rev   ..

 Align seg 1/1  to reverse of: US-08-541-033A-23  from: 1  to: 1506

     17 LeuGlyAlaLeuLeuThrIleValSerValValAlaAlaValPheThrGl 33
        |||::::::   |||:::   ::::::::::||||||   :::      ||
   1068 CTGGCAGTCCAGCTCCCAAGGCTTGCGGCGGTCGCCCACATA...CACGG 1022

     33 yGlyAlaSerLeuAlaLeuAlaAlaValGlyLeuAlaValMetValAlaA 50
        ||||:::::|||:::|||:::::::|||   :::|||::::::   :
   1021 CGGTGTCGCTCTTGTACTCGGAGATGCGGGCGCTGTTGTTCTTCTTCTTC 972

     50 spGluIleValLys.....AlaAlaThrGlyValSerPheIle 62
        ::   :::   :::      |||      :::|||:::::::::
    971 ATGTCCTGCACCGCCTGCAGCTGCTCGCGCGTGAAGCCGTTGGGCTCGTA 922

     63 GlnGlnAlaLeu.....AsnProIleMetGl 71
        :::   ||||||                  :::::::::::::
    921 CACGTAGCCCTGGGAGTCGGACAGCGACAGCACGATGGCGCCCTTCTCCA 872

     71 uHisValLeuLysProLeuMetGluLeuIleGlyLysAla.....I 85
        ::::   |||:::::::::::      ::::::::::|||
    871 GCAGCAGCTCCGCGCAGTACTGGGCCACGTTGCCCGCGCCAGACACCAGG 822

     85 leThrLysAlaLeuGluGlyLeuGlyVal..... 94
          :::   |||||||||:::|||::::::
    821 CAGCGCTTGCCCTTGAGGCTCTCGCCCTTGTCCTTCAGCACGTTCTCCAC 772

     95 ...Asp..ArgLysArgGlnArgTrpPro.AlaAlaLeuLeuValArgLe 109
           :::  |||:::|||   ||||||||| :::|||   |||
    771 AAACAGCACGGCGCCGTAGCCGGTGGCCTCGGGGCGGATCTC........ 730

    109 uSerProLeuCysHisGlyAspAlaValIleValValValAlaValValG 126
                              ::::::   :::::::::|||:::   |
    729 .....................GGAGCCGCCATACTCCTGGCCCTTCGGG 702

    126 lyLysGlyAlaAlaAlaLysLeuGlyAsnAlaLeuSerLysMetMetGly 142
        ||:::   |||:::::::   ||||||:::|||||||      :::::::::
    701 GTCAGCACGCCGGTGTAGTTCTTGGTGATGCGCTT......GTACTGGCC 658

    143 GluThrIleLysLysLeuValProAsnValLeuLysGlnLeuAlaGlnAs 159
        |||::::::   :::|||
    657 GAAAAGGTAGCCAATCTC.............................. 640

    159 nGlySerLysLeuPheThrGlnGlyMetGlnArgIleThrSerGlyLeuG 176
                                                  :::   :::   :
    639 ....................................GCGCGCGCCCACG 627

    176 lyAsnValGlySerLysMetGlyLeuGlnThrAsnAlaLeuSerLysGlu 192
```

```
      :::::|||::::::          |||:::::::::::::::::::   :::
    626 CCGATGTCGCCGGCGGGCACGTCCTGCACGTAGCTGATGTGGCGCTGCAG 577

      193 LeuValGlyAsnThrLeuAsn.............LysValAlaLeuGl 204
          |||      :::   |||:::                :::||||||||::
      576 CTCGGTCATGAAGGACTGGCAGAAGCGCATCACCTCCGCGTCGCTCTTGC 527

      204 yMetGluValThrAsnThrAlaAlaGlnSerAlaGly...GlyValAlaG 220
          ::::   |||         ||||||   :::|||:::   |||   :::
      526 CCTTGGGGTCGAAGTCGGAGCCGCCCTTGCCGCCGCCCATGGGCAGGGTG 477

      220 luGly.ValPheIleLysAsnAlaSerGluAlaLeuAlaAspPheMetLe 236
              :::|||:::|||      |||   |||      :::||||||::
      476 GTCAGGCTGTTCTTGAAGATCTGCTCAAAGGCA...AGGAACTTCATGAT 430

      236 uAlaArgPheAlaMetAspGlnIleGlnGlnTrpLeuLysGlnSer 251
          ::::|||||||:::          :::   |||   :::::::::
      429 GGACAGGTTCACG............GAGGGGTGGAAGCGCAGGCCG 396

seq_name: /cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-541-033A-7

seq_documentation_block:
; Sequence 7, Application US/08541033A
; Patent No. 5879941
;  GENERAL INFORMATION:
;    APPLICANT:  Schmidt, Robert R.
;    APPLICANT:  Miller, Philip
;    TITLE OF INVENTION:  NOVEL POLYPEPTIDES AND POLYNUCLEOTIDES
;    TITLE OF INVENTION:  RELATING TO THE '- AND  -SUBUNITS OF GLUTAMATE
;    TITLE OF INVENTION:  DEHYDROGENASES AND METHODS OF USE
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Saliwanchik & Saliwanchik
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  Florida
;      COUNTRY:  USA
;      ZIP:  32606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/541,033A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Whitlock, Ted W.
;      REGISTRATION NUMBER:  36,965
;      REFERENCE/DOCKET NUMBER:  UF155
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (904) 375-8100
;      TELEFAX:  (904) 372-5800
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1969 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-541-033A-7


alignment_scores:
            Quality:   91.00              Length:    267
              Ratio:  0.615                Gaps:     15
 Percent Similarity:  55.431   Percent Identity: 21.348

alignment_block:
 US-09-068-804-5 x US-08-541-033A-7/rev   ..

 Align seg 1/1  to reverse of: US-08-541-033A-7  from: 1  to: 1969

       17 LeuGlyAlaLeuLeuThrIleValSerValValAlaAlaValPheThrGl 33
          |||:::::   |||:::   :::::::::||||||   :::      ||
     1044 CTGGCAGTCCAGCTCCCAAGGCTTGCGGCGGTCGCCCACATA...CACGG 998

       33 yGlyAlaSerLeuAlaLeuAlaAlaValGlyLeuAlaValMetValAlaA 50
          ||||::::::|||:::|||:::::::::|||   :::|||::::::   :
      997 CGGTGTCGCTCTTGTACTCGGAGATGCGGGCGCTGTTGTICTTCTTCTTC 948

       50 spGluIleValLys............AlaAlaThrGlyValSerPheIle 62
          ::   :::   :::           |||      ::::|||:::::::::
      947 ATGTCCTGCACCGCCTGCAGCTGCTCGCGCGTGAAGCCGTTGGGCTCGTA 898

       63 GlnGlnAlaLeu.....................AsnProIleMetGl 71
          :::   ||||||                     :::::::::::::
      897 CACGTAGCCCTGGGAGTCGGACAGCGACAGCACGATGGCGCCCTTCTCCA 848

       71 uHisValLeuLysProLeuMetGluLeuIleGlyLysAla.....I 85
          ::::   |||:::::::::::::      ::::::::::|||
      847 GCAGCAGCTCCGCGCAGTACTGGGCCACGTTGCCCGCGCCAGACACCAGG 798

       85 leThrLysAlaLeuGluGlyLeuGlyVal..................... 94
            :::   ||||||||||:::|||::::::
      797 CAGCGCTTGCCCTTGAGGCTCTCGCCCTTGTCCTTCAGCACGTTCTCCAC 748

       95 ...Asp..ArgLysArgGlnArgTrpPro.AlaAlaLeuLeuValArgLe 109
             :::  |||:::|||   ||||||||| :::|||   |||
      747 AAACAGCACGGCGCCGTAGCCGGTGGCCTCGGGGCGGATCTC........ 706

      109 uSerProLeuCysHisGlyAspAlaValIleValValValAlaValValG 126
                                :::::   :::::::::|||:::   |
      705 .....................GGAGCCGCCATACTCCTGGCCCTTCGGG 678

      126 lyLysGlyAlaAlaAlaLysLeuGlyAsnAlaLeuSerLysMetMetGly 142
          ||:::   |||:::::::  |||||||:::|||||||      :::::::::
      677 GTCAGCACGCCGGTGTAGTTCTTGGTGATGCGCTT......GTACTGGCC 634

      143 GluThrIleLysLysLeuValProAsnValLeuLysGlnLeuAlaGlnAs 159
          |||:::::::   :::|||
      633 GAAAAGGTAGCCAATCTC............................... 616

      159 nGlySerLysLeuPheThrGlnGlyMetGlnArgIleThrSerGlyLeuG 176
                                                  :::   :::   :
      615 .....................................GCGCGCGCCCACG 603

      176 lyAsnValGlySerLysMetGlyLeuGlnThrAsnAlaLeuSerLysGlu 192
          :::::|||::::::          |||:::::::::::::::::::   :::
      602 CCGATGTCGCCGGCGGGCACGTCCTGCACGTAGCTGATGTGGCGCTGCAG 553

      193 LeuValGlyAsnThrLeuAsn.............LysValAlaLeuGl 204
          |||      :::   |||:::               :::||||||||::
      552 CTCGGTCATGAAGGACTGGCAGAAGCGCATCACCTCCGCGTCGCTCTTGC 503

      204 yMetGluValThrAsnThrAlaAlaGlnSerAlaGly...GlyValAlaG 220
          ::::   |||         ||||||   :::|||:::   |||   :::
      502 CCTTGGGGTCGAAGTCGGAGCCGCCCTTGCCGCCGCCCATGGGCAGGGTG 453

      220 luGly.ValPheIleLysAsnAlaSerGluAlaLeuAlaAspPheMetLe 236
              :::|||:::|||      |||   |||      :::||||||::
      452 GTCAGGCTGTTCTTGAAGATCTGCTCAAAGGCA...AGGAACTTCATGAT 406

      236 uAlaArgPheAlaMetAspGlnIleGlnGlnTrpLeuLysGlnSer 251
          ::::|||||||:::          :::   |||   :::::::::
      405 GGACAGGTTCACG............GAGGGGTGGAAGCGCAGGCCG 372

seq_name: /cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:PCT-US91-09422-18

seq_documentation_block:
; Sequence 18, Application PC/TUS9109422
;  GENERAL INFORMATION:
;    APPLICANT:  Mulvihill, Eileen R.
;    APPLICANT:  Hagen, Frederick S.
;    APPLICANT:  Houamed, Khaled M.
```

```
;    APPLICANT:  Almers, Wolfhard
;    TITLE OF INVENTION:  G PROTEIN-COUPLED GLUTAMATE RECEPTORS
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend
;      STREET:  One Market Plaza, Steuart Street Tower
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94105-1492
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US91/09422
;      FILING DATE:  19911212
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/672,007
;      FILING DATE:  18-MAR-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/648,481
;      FILING DATE:  30-JAN-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/626,806
;      FILING DATE:  12-DEC-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parmelee, Steven W.
;      REGISTRATION NUMBER:  31,990
;      REFERENCE/DOCKET NUMBER:  13952-6PC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 467-9600
;      TELEFAX:  (415) 543-5043
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4095 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    IMMEDIATE SOURCE:
;      CLONE:  SN30
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  463..3198
PCT-US91-09422-18


alignment_scores:
            Quality:   94.50              Length:    251
              Ratio:  0.836                Gaps:     11
 Percent Similarity:  45.020   Percent Identity: 23.506

alignment_block:
 US-09-068-804-5 x PCT-US91-09422-18/rev   ..

 Align seg 1/1  to reverse of: PCT-US91-09422-18  from: 1  to: 4095

       6 ThrAsnArgIleMetGlyCysIleGlyLysValLeuGlyAlaLeuLeuTh 22
         :::::::|||   :::   |||          |||:::||||||   |||||
    2181 GCAGCTCGTGCGGTTCTCTGTGGGCCGCATGTCGTAGGGGCAGGTCTTAC 2132

      22 rIleValSerValValAlaAlaValPheThrGlyGlyAlaSerLeuAlaL 39
         |   |||:::   |||      |||::::::   ||||||:::||||||:
    2131 AGGTGTAGCG...GTCCACTTGGTACTGGTACCCGGTGCAGGGCTCGCAG 2085

      39 euAlaAlaValGlyLeuAlaValMetValAlaAspGluIleValLysAla 55
         :::::|||::::::
    2084 TGCCAGCAGCAAGC..................................... 2071

      56 AlaThrGlyValSerPheIleGlnGlnAlaLeuAsnProIleMetGluHi 72
                                                          ||
    2070 .....................................................CA 2069

      72 sValLeuLysProLeuMetGluLeuIleGlyLysAlaIleThrLysAlaL 89
         |:::|||   :::|||:::   |||   |||                   |
    2068 TGCCCTTCACAGTCTTCTTTCGCTCCCCGGG.................C 2037

      89 euGluGlyLeuGlyValAspArgLysArgGln.............ArgTrp 101
         ||:::|||      :::|||      ||||||               |||
    2036 TGGCAGGGCAGACTGCAGATGGAGCGCGGCAGCTGCTGGCCACTCCCTGG 1987

     102 ProAlaAlaLeuLeuValArgLeuSerProLeuCysHisGlyAspAlaVa 118
         |||                  |||   ||||||:::
    1986 CCA.................CTGCATCCGCTCTATTCTGAG........ 1963

     118 lIleValValValAlaValValGlyLysGlyAlaAlaAlaLysLeuGlyA 135
               |||   :::|||      ::::::::::::      ||||||
    1962 .......GTGCAGGTGGTCTGTCCACGAGCCGATGACCTTGTACTCGGCC 1920

     135 snAlaLeuSerLysMetMet............................. 141
           |||:::::::::::::::
    1919 GAGCCATTGCGCAGTTGGTACTGGTAGATGTCGTAGCGCCCCGGTGCGTC 1870

     142 ..............GlyGluThrIleLysLysLeu.............. 148
                       |||      :::  :::
    1869 TCCGTTCTCATTGAAGGTTACAGGGTTCCCCGCAATGCCTGAGAAGTTGA 1820

     149 .ValProAsnValLeuLysGlnLeuAla..................... 157
          ||||||:::||||||||||||||||:::
    1819 CGTTCCTGATGTACTTAAGCAGCTGGGTGCCATCCACGGGGTCCATGCGA 1770

     158 ..............GlnAsnGlySerLysLeuPheThrGlnGlyMetGln 169
                        :::|||::::::::::   ::::::|||:::|||
    1769 GGGCAGAGTCCTACGCGGCCGGGACACAGGTCACGGTGCATGGCGTGCAG 1720

     170 ArgIleThrSerGlyLeuGlyAsnValGlySerLysMetGlyLeuGlnTh 186
         |||::::::   |||:::   :::|||:::      :::   |||:::
    1719 CGCGTGGCCCATGGCGTACACAGCGTCAATCACGAACTGCACCTTCCCCT 1670

     186 rAsnAlaLeuSerLysGluLeu............................ 193
               :::::::::   |||
    1669 CCTGCTCATAGGCCGAGTCCTGCCCGATGCGCTCTCGGTTGGTGCACTTC 1620

     194 .....ValGlyAsnThrLeuAsnLysValAla......LeuGlyMetGlu 206
              |||::::::   |||:::::::||||||      |||:::||||||
    1619 TTGATGTGGCTTCCCTTCTTGAGCGCGTGGCGGCTCAACTTGCAATGGAA 1570

     207 Val 207
         |||
    1569 GTT 1567

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-325-267A-1

seq_documentation_block:
; Sequence 1, Application US/08325267A
; Patent No. 5585271
;  GENERAL INFORMATION:
;    APPLICANT:  WATARI, JUNJI
;    APPLICANT:  TAKATA, YOSHIHIRO
;    APPLICANT:  OGAWA, MASAHIRO
;    APPLICANT:  PENTTILA, MERJA
;    APPLICANT:  ONNELA, MAIJA-LEENA
;    APPLICANT:  KERANEN, SIRKKA
;    TITLE OF INVENTION:  YEAST AGGLUTINATION GENES AND YEAST
;    TITLE OF INVENTION:  CONTAINING THEM
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT
;      STREET:  1755 S. JEFFERSON DAVIS HWY, SUITE 400
;      CITY:  ARLINGTON
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22202
```

```
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/325,267A
;      FILING DATE:  18-NOV-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP PCT/JP94/00290
;      FILING DATE:  24-FEB-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 38871/1993
;      FILING DATE:  26-FEB-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OBLON, NORMAN F
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  2589-023-0XPCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-413-3000
;      TELEFAX:  703-413-2220
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4614 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Saccharomyces cerevisiae
;      STRAIN:  ABXL-1D
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..4614
US-08-325-267A-1


alignment_scores:
            Quality:   95.00            Length:    194
              Ratio:  0.841              Gaps:      6
 Percent Similarity:  58.247   Percent Identity: 23.196

alignment_block:
 US-09-068-804-5 x US-08-325-267A-1/rev   ..

 Align seg 1/1  to reverse of: US-08-325-267A-1  from: 1  to: 4614

      16 ValLeuGlyAlaLeuLeuThrIleValSerValValAlaAlaValPheTh 32
         |||:::   |||:::::::::::::|||:::|||:::   ::::::   ::
    1376 GTAGTCATAGCAGTAGTGGCTGTTGTTGGTGTTCTGATGACAATGATGGT 1327

      32 rGlyGlyAlaSerLeuAlaLeuAlaAlaValGlyLeuAlaValMetValA 49
         ::::::::::::   :::|||:::::::|||   :::::::::   |||:
    1326 TTCATCAGTTGGCAAACCGTTGGTACCGGTGACGGTGGTCATTTCAGTGG 1277

      49 laAspGluIleValLysAlaAlaThrGlyValSerPheIleGlnGlnAla 65
         ::   |||:::                  |||||||||         :::
    1276 ATGTAGAGGTA..............AAAGTGTCGTTCCATGGCTGAGTT 1242

      66 LeuAsnProIleMetGluHisValLeuLysProLeuMetGluLeuIleGl 82
         :::         :::      |||:::      ||||||   ::::::
    1241 GTAGTCATGGCAGTAGTGGCTGTTGTTGGTGTTCTGATGACAATGATGGT 1192

      82 yLysAlaIleThrLysAlaLeuGluGlyLeuGlyValAspArgLysArgG 99
            ::::::   |||:::|||      :::   |||
    1191 CTCATCAGTTGGCAAACCATTGGTACCGGTGACTGTG............. 1155

      99 lnArgTrpProAlaAlaLeuLeuValArgLeuSerProLeuCysHisGly 115
                ::::::::   |||   :::      |||   ||||||
    1154 ........GTCAATTCGGTAGAAGTAGAGGTAAAAGTGCTGTTCCATGGC 1113

      116 AspAlaValIleVal...ValValAlaValValGlyLysGlyAlaAlaAl 131
          ::::::|||::::::   ||||||||||||||||||
     1112 TCAGTTGTAGTTATGGCAGTAGTGGCTGTTGTTGGT.............. 1077

      131 aLysLeuGlyAsnAlaLeuSerLysMetMetGlyGluThrIleLysLys. 147
                    :::|||   :::||||||      ::::::::::::
     1076 .............GTTCTGATTACAATGATGGTTTCGTCAGTTCGTACAC 1040

      148 ..LeuValProAsnValLeuLysGlnLeuAlaGlnAsnGlySerLysLeu 163
            |||||||||   ::::::   :::   ::::::         |||:::
     1039 CATTGGTGCCAGTGACTGTGGTCAATTCGGTAGAAGTAGAGGTAAAAGTG 990

      164 PheThrGlnGlyMetGlnArgIleThrSerGlyLeuGlyAsnValGlySe 180
          :::   :::|||         :::         |||:::   ||||||
      989 CTGTTCCATGGCTCAGTTGTAGTTATGATGGTGCTAGCAGTTGTTGGAGT 940

      180 rLysMetGlyLeuGlnThrAsnAlaLeuSerLysGluLeuValGlyAsnT 197
             |||   :::   :::::::::::::::::   ||||||      |
      939 TCTGATGACAATGACGGTTTCGTCAGTTGGAACGCCGTTGGTACCGGTGA 890

      197 hrLeuAsnLysValAlaLeuGlyMetGluVal 207
          ||         |||:::::::::::::::|||
      889 CG.........GTGGTCATTTCAGTAGATGTA 867

seq_name: /cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-541-033A-19

seq_documentation_block:
; Sequence 19, Application US/08541033A
; Patent No. 5879941
;  GENERAL INFORMATION:
;    APPLICANT:  Schmidt, Robert R.
;    APPLICANT:  Miller, Philip
;    TITLE OF INVENTION:  NOVEL POLYPEPTIDES AND POLYNUCLEOTIDES
;    TITLE OF INVENTION:  RELATING TO THE '- AND  -SUBUNITS OF GLUTAMATE
;    TITLE OF INVENTION:  DEHYDROGENASES AND METHODS OF USE
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Saliwanchik & Saliwanchik
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  Florida
;      COUNTRY:  USA
;      ZIP:  32606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/541,033A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Whitlock, Ted W.
;      REGISTRATION NUMBER:  36,965
;      REFERENCE/DOCKET NUMBER:  UF155
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (904) 375-8100
;      TELEFAX:  (904) 372-5800
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2096 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  CDNA
US-08-541-033A-19


alignment_scores:
            Quality:   91.00            Length:    267
              Ratio:  0.615              Gaps:      15
 Percent Similarity:  55.431   Percent Identity: 21.348
```

```
alignment_block:
 US-09-068-804-5 x US-08-541-033A-19/rev   ..

 Align seg 1/1  to reverse of: US-08-541-033A-19  from: 1  to: 2096

      17 LeuGlyAlaLeuLeuThrIleValSerValValAlaAlaValPheThrGl 33
         |||::::::   |||:::   :::::::::||||||   :::      ||
    1169 CTGGCAGTCCAGCTCCCAAGGCTTGCGGCGGTCGCCCACATA...CACGG 1123

      33 yGlyAlaSerLeuAlaLeuAlaAlaValGlyLeuAlaValMetValAlaA 50
         ||||::::::|||:::|||::::::::::|||   :::|||::::::   :
    1122 CGGTGTCGCTCTTGTACTCGGAGATGCGGGCGCTGTTGTTCTTCTTCTTC 1073

      50 spGluIleValLys............AlaAlaThrGlyValSerPheIle 62
         ::   :::   :::            |||      :::|||::::::::::
    1072 ATGTCCTGCACCGCCTGCAGCTGCTCGCGCGTGAAGCCGTTGGGCTCGTA 1023

      63 GlnGlnAlaLeu.....................AsnProIleMetGl 71
         :::   ||||||                     ::::::::::::::
    1022 CACGTAGCCCTGGGAGTCGGACAGCGACAGCACGATGGCGCCCTTCTCCA 973

      71 uHisValLeuLysProLeuMetGluLeuIleGlyLysAla.........I 85
         ::::   |||:::::::::::::      :::::::::|||
     972 GCAGCAGCTCCGCGCAGTACTGGGCCACGTTGCCCGCGCCAGACACCAGG 923

      85 leThrLysAlaLeuGluGlyLeuGlyVal..................... 94
           :::   |||||||||:::|||::::::
     922 CAGCGCTTGCCCTTGAGGCTCTCGCCCTTGTCCTTCAGCACGTTCTCCAC 873

      95 ...Asp..ArgLysArgGlnArgTrpPro.AlaAlaLeuLeuValArgLe 109
            :::  |||:::|||   |||||||||| :::|||   |||
     872 AAACAGCACGGCGCCGTAGCCGGTGGCCTCGGGGCGGATCTC........ 831

     109 uSerProLeuCysHisGlyAspAlaValIleValValValAlaValValG 126
                          ::::::   :::::::::|||:::   |
     830 .....................GGAGCCGCCATACTCCTGGCCCTTCGGG 803

     126 lyLysGlyAlaAlaAlaLysLeuGlyAsnAlaLeuSerLysMetMetGly 142
         ||:::   |||::::::   ||||||:::||||||      :::::::::
     802 GTCAGCACGCCGGTGTAGTTCTTGGTGATGCGCTT......GTACTGGCC 759

     143 GluThrIleLysLysLeuValProAsnValLeuLysGlnLeuAlaGlnAs 159
         |||::::::   :::|||
     758 GAAAAGGTAGCCAATCTC............................... 741

     159 nGlySerLysLeuPheThrGlnGlyMetGlnArgIleThrSerGlyLeuG 176
                                                :::   :::   :
     740 ....................................GCGCGCGCCCACG 728

     176 lyAsnValGlySerLysMetGlyLeuGlnThrAsnAlaLeuSerLysGlu 192
         :::::|||::::::       |||:::::::::::::::::::   :::
     727 CCGATGTCGCCGGCGGGCACGTCCTGCACGTAGCTGATGTGGCGCTGCAG 678

     193 LeuValGlyAsnThrLeuAsn..............LysValAlaLeuGl 204
         |||       :::   |||:::            :::||||||||::
     677 CTCGGTCATGAAGGACTGGCAGAAGCGCATCACCTCCGCGTCGCTCTTGC 628

     204 yMetGluValThrAsnThrAlaAlaGlnSerAlaGly...GlyValAlaG 220
         ::::   |||         ||||||   :::|||:::   |||   :::
     627 CCTTGGGGTCGAAGTCGGAGCCGCCCTTGCCGCCGCCCATGGGCAGGGTG 578

     220 luGly.ValPheIleLysAsnAlaSerGluAlaLeuAlaAspPheMetLe 236
              :::|||:::|||     |||   |||      :::|||||||::
     577 GTCAGGCTGTTCTTGAAGATCTGCTCAAAGGCA...AGGAACTTCATGAT 531

     236 uAlaArgPheAlaMetAspGlnIleGlnGlnTrpLeuLysGlnSer 251
         ::::||||||:::            :::   |||   :::::::::
     530 GGACAGGTTCACG............GAGGGGTGGAAGCGCAGGCCG 497

 seq_name: /cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-541-033A-3

 seq_documentation_block:

; Sequence 3, Application US/08541033A
; Patent No. 5879941
;  GENERAL INFORMATION:
;    APPLICANT:  Schmidt, Robert R.
;    APPLICANT:  Miller, Philip
;    TITLE OF INVENTION:  NOVEL POLYPEPTIDES AND POLYNUCLEOTIDES
;    TITLE OF INVENTION:  RELATING TO THE '- AND  -SUBUNITS OF GLUTAMATE
;    TITLE OF INVENTION:  DEHYDROGENASES AND METHODS OF USE
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Saliwanchik & Saliwanchik
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  Florida
;      COUNTRY:  USA
;      ZIP:  32606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/541,033A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Whitlock, Ted W.
;      REGISTRATION NUMBER:  36,965
;      REFERENCE/DOCKET NUMBER:  UF155
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (904) 375-8100
;      TELEFAX:  (904) 372-5800
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2099 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  33..1568
US-08-541-033A-3


alignment_scores:
            Quality:   91.00            Length:    267
              Ratio:  0.615              Gaps:     15
 Percent Similarity:  55.431   Percent Identity: 21.348

alignment_block:
 US-09-068-804-5 x US-08-541-033A-3/rev   ..

 Align seg 1/1  to reverse of: US-08-541-033A-3  from: 1  to: 2099

      17 LeuGlyAlaLeuLeuThrIleValSerValValAlaAlaValPheThrGl 33
         |||::::::   |||:::   :::::::::||||||   :::      ||
    1172 CTGGCAGTCCAGCTCCCAAGGCTTGCGGCGGTCGCCCACATA...CACGG 1126

      33 yGlyAlaSerLeuAlaLeuAlaAlaValGlyLeuAlaValMetValAlaA 50
         ||||::::::|||:::|||::::::::::|||   :::|||::::::   :
    1125 CGGTGTCGCTCTTGTACTCGGAGATGCGGGCGCTGTTGTTCTTCTTCTTC 1076

      50 spGluIleValLys............AlaAlaThrGlyValSerPheIle 62
         ::   :::   :::            |||      :::|||::::::::::
    1075 ATGTCCTGCACCGCCTGCAGCTGCTCGCGCGTGAAGCCGTTGGGCTCGTA 1026

      63 GlnGlnAlaLeu.....................AsnProIleMetGl 71
         :::   ||||||                     ::::::::::::::
    1025 CACGTAGCCCTGGGAGTCGGACAGCGACAGCACGATGGCGCCCTTCTCCA 976

      71 uHisValLeuLysProLeuMetGluLeuIleGlyLysAla.........I 85
         ::::   |||:::::::::::::      :::::::::|||
```

```
     975 GCAGCAGCTCCGCGCAGTACTGGGCCACGTTGCCCGCGCCAGACACCAGG 926

      85 leThrLysAlaLeuGluGlyLeuGlyVal..................... 94
         :::   |||||||||:::|||:::::
     925 CAGCGCTTGCCCTTGAGGCTCTCGCCCTTGTCCTTCAGCACGTTCTCCAC 876

      95 ...Asp..ArgLysArgGlnArgTrpPro.AlaAlaLeuLeuValArgLe 109
            :::  |||:::|||   |||||||||| :::|||   |||
     875 AAACAGCACGGCGCCGTAGCCGGTGGCCTCGGGGCGGATCTC........ 834

     109 uSerProLeuCysHisGlyAspAlaValIleValValValAlaValValG 126
                           ::::::   :::::::::|||:::   |
     833 .....................GGAGCCGCCATACTCCTGGCCCTTCGGG 806

     126 lyLysGlyAlaAlaAlaLysLeuGlyAsnAlaLeuSerLysMetMetGly 142
         ||:::   |||::::::   ||||||:::|||||||      :::::::::
     805 GTCAGCACGCCGGTGTAGTTCTTGGTGATGCGCTT......GTACTGGCC 762

     143 GluThrIleLysLysLeuValProAsnValLeuLysGlnLeuAlaGlnAs 159
         |||::::::   :::|||
     761 GAAAAGGTAGCCAATCTC............................... 744

     159 nGlySerLysLeuPheThrGlnGlyMetGlnArgIleThrSerGlyLeuG 176
                                          :::   :::   :
     743 ....................................GCGCGCGCCCACG 731

     176 lyAsnValGlySerLysMetGlyLeuGlnThrAsnAlaLeuSerLysGlu 192
         :::::|||::::::      |||:::::::::::::::::::   :::
     730 CCGATGTCGCCGGCGGGCACGTCCTGCACGTAGCTGATGTGGCGCTGCAG 681

     193 LeuValGlyAsnThrLeuAsn..............LysValAlaLeuGl 204
         |||      :::   |||:::             :::||||||||::
     680 CTCGGTCATGAAGGACTGGCAGAAGCGCATCACCTCCGCGTCGCTCTTGC 631

     204 yMetGluValThrAsnThrAlaAlaGlnSerAlaGly...GlyValAlaG 220
         ::::   |||         ||||||   :::|||:::   |||   :::
     630 CCTTGGGGTCGAAGTCGGAGCCGCCCTTGCCGCCGCCCATGGGCAGGGTG 581

     220 luGly.ValPheIleLysAsnAlaSerGluAlaLeuAlaAspPheMetLe 236
            :::|||:::|||     |||   |||      :::|||||||::
     580 GTCAGGCTGTTCTTGAAGATCTGCTCAAAGGCA...AGGAACTTCATGAT 534

     236 uAlaArgPheAlaMetAspGlnIleGlnGlnTrpLeuLysGlnSer 251
         ::::|||||||:::            :::   |||   :::::::::
     533 GGACAGGTTCACG............GAGGGGTGGAAGCGCAGGCCG 500
```

seq_name: /cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-541-033A-18

```
seq_documentation_block:
; Sequence 18, Application US/08541033A
; Patent No. 5879941
;  GENERAL INFORMATION:
;    APPLICANT:  Schmidt, Robert R.
;    APPLICANT:  Miller, Philip
;    TITLE OF INVENTION:  NOVEL POLYPEPTIDES AND POLYNUCLEOTIDES
;    TITLE OF INVENTION:  RELATING TO THE `- AND  -SUBUNITS OF GLUTAMATE
;    TITLE OF INVENTION:  DEHYDROGENASES AND METHODS OF USE
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Saliwanchik & Saliwanchik
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  Florida
;      COUNTRY:  USA
;      ZIP:  32606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/541,033A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Whitlock, Ted W.
;      REGISTRATION NUMBER:  36,965
;      REFERENCE/DOCKET NUMBER:  UF155
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (904) 375-8100
;      TELEFAX:  (904) 372-5800
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2137 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  CDNA
US-08-541-033A-18


alignment_scores:
            Quality:   91.00              Length:    267
              Ratio:  0.615                Gaps:     15
 Percent Similarity:  55.431   Percent Identity: 21.348

alignment_block:
 US-09-068-804-5 x US-08-541-033A-18/rev   ..

 Align seg 1/1  to reverse of: US-08-541-033A-18  from: 1  to: 2137

      17 LeuGlyAlaLeuLeuThrIleValSerValValAlaAlaValPheThrGl 33
         |||:::::::   |||:::   ::::::::::||||||   :::      ||
    1211 CTGGCAGTCCAGCTCCCAAGGCTTGCGGCGGTCGCCCACATA...CACGG 1165

      33 yGlyAlaSerLeuAlaLeuAlaAlaValGlyLeuAlaValMetValAlaA 50
         ||||::::::|||:::|||::::::::::|||   :::|||::::::   :
    1164 CGGTGTCGCTCTTGTACTCGGAGATGCGGGCGCTGTTGTTCTTCTTCTTC 1115

      50 spGluIleValLys............AlaAlaThrGlyValSerPheIle 62
         ::   :::   :::             |||      :::|||::::::::
    1114 ATGTCCTGCACCGCCTGCAGCTGCTCGCGCGTGAAGCCGTTGGGCTCGTA 1065

      63 GlnGlnAlaLeu.......................AsnProIleMetGl 71
         :::   ||||||                      ::::::::::::::
    1064 CACGTAGCCCTGGGAGTCGGACAGCGACAGCACGATGGCGCCCTTCTCCA 1015

      71 uHisValLeuLysProLeuMetGluLeuIleGlyLysAla.........I 85
         ::::   |||:::::::::::::      :::::::::|||
    1014 GCAGCAGCTCCGCGCAGTACTGGGCCACGTTGCCCGCGCCAGACACCAGG 965

      85 leThrLysAlaLeuGluGlyLeuGlyVal..................... 94
         :::   ||||||||||:::|||:::::
     964 CAGCGCTTGCCCTTGAGGCTCTCGCCCTTGTCCTTCAGCACGTTCTCCAC 915

      95 ...Asp..ArgLysArgGlnArgTrpPro.AlaAlaLeuLeuValArgLe 109
            :::  |||:::|||   |||||||||| :::|||   |||
     914 AAACAGCACGGCGCCGTAGCCGGTGGCCTCGGGGCGGATCTC........ 873

     109 uSerProLeuCysHisGlyAspAlaValIleValValValAlaValValG 126
                           ::::::   :::::::::|||:::   |
     872 .....................GGAGCCGCCATACTCCTGGCCCTTCGGG 845

     126 lyLysGlyAlaAlaAlaLysLeuGlyAsnAlaLeuSerLysMetMetGly 142
         ||:::   |||::::::   ||||||:::|||||||      :::::::::
     844 GTCAGCACGCCGGTGTAGTTCTTGGTGATGCGCTT......GTACTGGCC 801

     143 GluThrIleLysLysLeuValProAsnValLeuLysGlnLeuAlaGlnAs 159
         |||::::::   :::|||
     800 GAAAAGGTAGCCAATCTC............................... 783

     159 nGlySerLysLeuPheThrGlnGlyMetGlnArgIleThrSerGlyLeuG 176
                                          :::   :::   :
     782 ....................................GCGCGCGCCCACG 770

     176 lyAsnValGlySerLysMetGlyLeuGlnThrAsnAlaLeuSerLysGlu 192
```

```
      :::::|||::::::          |||:::::::::::::::::::   :::
    769 CCGATGTCGCCGGCGGGCACGTCCTGCACGTAGCTGATGTGGCGCTGCAG 720

    193 LeuValGlyAsnThrLeuAsn..............LysValAlaLeuGl 204
        |||       :::   |||:::              :::||||||||||::
    719 CTCGGTCATGAAGGACTGGCAGAAGCGCATCACCTCCGCGTCGCTCTTGC 670

    204 yMetGluValThrAsnThrAlaAlaGlnSerAlaGly...GlyValAlaG 220
        ::::   |||          ||||||   :::|||:::   |||   :::
    669 CCTTGGGGTCGAAGTCGGAGCCGCCCTTGCCGCCGCCCATGGGCAGGGTG 620

    220 luGly.ValPheIleLysAsnAlaSerGluAlaLeuAlaAspPheMetLe 236
              :::|||:::|||      |||   |||      :::|||||||::
    619 GTCAGGCTGTTCTTGAAGATCTGCTCAAAGGCA...AGGAACTTCATGAT 573

    236 uAlaArgPheAlaMetAspGlnIleGlnGlnTrpLeuLysGlnSer 251
        ::::|||||::            :::   |||   :::::::::
    572 GGACAGGTTCACG............GAGGGGTGGAAGCGCAGGCCG 539

seq_name: /cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-541-033A-1

seq_documentation_block:
; Sequence 1, Application US/08541033A
; Patent No. 5879941
;  GENERAL INFORMATION:
;    APPLICANT:  Schmidt, Robert R.
;    APPLICANT:  Miller, Philip
;    TITLE OF INVENTION:  NOVEL POLYPEPTIDES AND POLYNUCLEOTIDES
;    TITLE OF INVENTION:  RELATING TO THE `- AND  -SUBUNITS OF GLUTAMATE
;    TITLE OF INVENTION:  DEHYDROGENASES AND METHODS OF USE
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Saliwanchik & Saliwanchik
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  Florida
;      COUNTRY:  USA
;      ZIP:  32606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/541,033A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Whitlock, Ted W.
;      REGISTRATION NUMBER:  36,965
;      REFERENCE/DOCKET NUMBER:  UF155
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (904) 375-8100
;      TELEFAX:  (904) 372-5800
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2140 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  33..1610
US-08-541-033A-1


alignment_scores:
           Quality:   91.00            Length:    267
             Ratio:  0.615              Gaps:     15
 Percent Similarity:  55.431   Percent Identity: 21.348

alignment_block:
```

```
US-09-068-804-5 x US-08-541-033A-1/rev   ..

Align seg 1/1  to reverse of: US-08-541-033A-1  from: 1  to: 2140

     17 LeuGlyAlaLeuLeuThrIleValSerValValAlaAlaValPheThrGl 33
        |||::::::   |||:::   :::::::::||||||   :::      ||
   1214 CTGGCAGTCCAGCTCCCAAGGCTTGCGGCGGTCGCCCACATA...CACGG 1168

     33 yGlyAlaSerLeuAlaLeuAlaAlaValGlyLeuAlaValMetValAlaA 50
        ||||:::::|||:::|||:::::::::|||   :::|||::::::   :
   1167 CGGTGTCGCTCTTGTACTCGGAGATGCGGGCGCTGTTGTTCTTCTTCTTC 1118

     50 spGluIleValLys............AlaAlaThrGlyValSerPheIle 62
        ::   :::   :::            |||      :::|||:::::::::
   1117 ATGTCCTGCACCGCCTGCAGCTGCTCGCGCGTGAAGCCGTTGGGCTCGTA 1068

     63 GlnGlnAlaLeu.....................AsnProIleMetGl 71
        :::   ||||||                     :::::::::::::::
   1067 CACGTAGCCCTGGGAGTCGGACAGCGACAGCACGATGGCGCCCTTCTCCA 1018

     71 uHisValLeuLysProLeuMetGluLeuIleGlyLysAla.........I 85
        ::::   |||:::::::::::::      ::::::::::|||
   1017 GCAGCAGCTCCGCGCAGTACTGGGCCACGTTGCCCGCGCCAGACACCAGG 968

     85 leThrLysAlaLeuGluGlyLeuGlyVal..................... 94
          :::   ||||||||||:::|||::::::
    967 CAGCGCTTGCCCTTGAGGCTCTCGCCCTTGTCCTTCAGCACGTTCTCCAC 918

     95 ...Asp..ArgLysArgGlnArgTrpPro.AlaAlaLeuLeuValArgLe 109
           :::  |||:::|||   ||||||||| :::|||   |||
    917 AAACAGCACGGCGCCGTAGCCGGTGGCCTCGGGGCGGATCTC........ 876

    109 uSerProLeuCysHisGlyAspAlaValIleValValValAlaValValG 126
                              ::::::   :::::::::|||:::   |
    875 .....................GGAGCCGCCATACTCCTGGCCCTTCGGG 848

    126 lyLysGlyAlaAlaAlaLysLeuGlyAsnAlaLeuSerLysMetMetGly 142
        ||:::   |||::::::   ||||||:::|||||||      :::::::::
    847 GTCAGCACGCCGGTGTAGTTCTTGGTGATGCGCTT......GTACTGGCC 804

    143 GluThrIleLysLysLeuValProAsnValLeuLysGlnLeuAlaGlnAs 159
        |||::::::   :::|||
    803 GAAAAGGTAGCCAATCTC............................... 786

    159 nGlySerLysLeuPheThrGlnGlyMetGlnArgIleThrSerGlyLeuG 176
                                                :::   :::   :
    785 ..................................GCGCGCGCCCACG 773

    176 lyAsnValGlySerLysMetGlyLeuGlnThrAsnAlaLeuSerLysGlu 192
        ::::::|||::::::         |||::::::::::::::::::   :::
    772 CCGATGTCGCCGGCGGGCACGTCCTGCACGTAGCTGATGTGGCGCTGCAG 723

    193 LeuValGlyAsnThrLeuAsn..............LysValAlaLeuGl 204
        |||       :::   |||:::              :::||||||||||::
    722 CTCGGTCATGAAGGACTGGCAGAAGCGCATCACCTCCGCGTCGCTCTTGC 673

    204 yMetGluValThrAsnThrAlaAlaGlnSerAlaGly...GlyValAlaG 220
        ::::   |||          ||||||   :::|||:::   |||   :::
    672 CCTTGGGGTCGAAGTCGGAGCCGCCCTTGCCGCCGCCCATGGGCAGGGTG 623

    220 luGly.ValPheIleLysAsnAlaSerGluAlaLeuAlaAspPheMetLe 236
              :::|||:::|||      |||   |||      :::|||||||::
    622 GTCAGGCTGTTCTTGAAGATCTGCTCAAAGGCA...AGGAACTTCATGAT 576

    236 uAlaArgPheAlaMetAspGlnIleGlnGlnTrpLeuLysGlnSer 251
        ::::|||||::            :::   |||   :::::::::
    575 GGACAGGTTCACG............GAGGGGTGGAAGCGCAGGCCG 542

seq_name: /cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:PCT-US93-10500-1

seq_documentation_block:
; Sequence 1, Application PC/TUS9310500
;  GENERAL INFORMATION:
```

```
;    APPLICANT:  Chang, Yung-Fu
;    TITLE OF INVENTION:  Recombinant Vaccine For Procine
;    TITLE OF INVENTION:  Pleuropneumoniae
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Alan S. Korman
;      STREET:  1600 Empire Tower
;      CITY:  Buffalo
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  14202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/10500
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/972,229
;      FILING DATE:  05-NOV-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Korman, Alan S.
;      REGISTRATION NUMBER:  33,932
;      REFERENCE/DOCKET NUMBER:  19603/00001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  716-853-8104
;      TELEFAX:  716-853-8109
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3828 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:  N-terminal
;    ORIGINAL SOURCE:
;      ORGANISM:  Actinobacillus - pleuropneumonia
;      STRAIN:  Serotypes 2, 3, 4, 6 and 8
;      INDIVIDUAL ISOLATE:  swine
;      CELL TYPE:  gram negative bacterium
;    IMMEDIATE SOURCE:
;      LIBRARY:  A. pleuropneumoniae DNA in Lambda - dash
;      CLONE:  (Lambda) yfc 26-28 and (Lambda) yfc 31-32
PCT-US93-10500-1


alignment_scores:
            Quality:   92.00              Length:    211
              Ratio:  0.724                Gaps:     10
 Percent Similarity:  60.190   Percent Identity: 22.749

alignment_block:
 US-09-068-804-5 x PCT-US93-10500-1   ..

 Align seg 1/1  to: PCT-US93-10500-1  from: 1  to: 3828

       9 IleMetGlyCysIleGlyLysValLeuGlyAlaLeuLeuThrIleValSe 25
         |||::::::  |||     |||||||||:::::::|||:::   :::::
    1019 ATTATTTCAGGAATTCAATCTGTATTAGGTACTGTATTAGCAGGTATTAA 1068

      25 rValValAlaAlaValPheThrGlyGlyAlaSerLeuAlaLeu......A 40
         ::::   ::::|||:::::::::|||||||:::   |||:::|||     |
    1069 TCTTAATGAAGCTATTATTAGTGGCGGTTCAGAGCTCGAATTAGCTGAAG 1118

      40 laAlaValGlyLeuAlaValMetValAlaAspGluIleValLysAlaAla 56
         ||:::|||:::|||||      :::::::::::|||:::|||:::::
    1119 CTGGTGTTTCTTTAGCCTCTGAGCTCGTTAGCAATATTGCTAAAGGTACA 1168

      57 ThrGlyVal...SerPheIleGln......GlnAlaLeuAsnProIleMe 70
         |||   :::   :::|||:::          |||:::::::::   :::
    1169 ACAACAATAGATGCTTTCACTACACAAATCCAGAACTTTGGGAAATTAGC 1218

      70 tGluHis.................ValLeuLysProLeuMetGluLeuI 81
         |||:::                  |||   :::::::|||   ::::::
    1219 GGAAAATGCTAAAGGGTTAGGTGGTGTTGGCCGCCAATTACAGAATATTT 1268

      81 leGlyLysAlaIleThrLysAlaLeuGluGlyLeuGlyValAspArgLys 97
           |||   |||::::::|||:::      |||||||||:::|||
    1269 CAGGTTCTGCATTAAGCAAAACT......GGATTAGGTTTGGATATTATC 1312

      98 ArgGlnArgTrpProAlaAlaLeuLeu.........ValArgLeuSerPr 111
                    ::::::||||||         :::|||      ::
    1313 ..............TCAAGCTTACTTTCAGGAGTAACTCGAAGTTTTGC 1347

     111 oLeuCysHisGlyAspAlaValIleValValValAlaValValGlyLysG 128
         :|||   :::   :::|||                              :
    1348 TTTACGGAATAAGAATGCT.............................T 1367

     128 lyAlaAlaAlaLysLeuGlyAsnAlaLeu.........SerLysMetMet 141
         ::::::::::|||:::::::::::::::         ::::::::::::
    1368 CAACAAGCACTAAAGTTGCTGCTGGCTTTGAACTCTCAAATCAAGTAATT 1417

     142 GlyGluThrIleLysLysLeuValProAsnValLeuLysGlnLeuAlaGl 158
         |||   ::::::|||   :::   :::   :::|||
    1418 GGTGGTATTACGAAAGCAGTATCAAGCTATATTCTT.............. 1453

     158 nAsnGlySerLysLeuPheThrGlnGlyMetGlnArgIleThrSerGlyL 175
                                      ||||||:::   :::||||
    1454 ..........................GCACAGCGTTTACGTGCTGGTT 1475

     175 euGlyAsnValGlySerLysMetGlyLeuGlnThrAsnAlaLeuSerLys 191
         ||::::::::::|||:::       :::|||   ::::::::::::|||
    1476 TATCAACGACAGGTCCTGCTGCAGCACTAATTGCGTCTAGTATTTCTTTA 1525

     192 GluLeuValGlyAsnThrLeuAsnLysValAla 202
         ::::::       ::::::   :::||||||
    1526 GCAATCAGTCCATTGGCGTTTTTACGTGTAGCT 1558

seq_name: /cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:PCT-US91-01327-12

seq_documentation_block:
; Sequence 12, Application PC/TUS9101327
;  GENERAL INFORMATION:
;    APPLICANT:  Draper et al.
;    TITLE OF INVENTION:  Oligonucleotide Therapies for
;    TITLE OF INVENTION:  Modulating the Effects of Herpesviruses
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock Washburn Kurtz
;      ADDRESSEE:  Mackiewicz & Norris
;      STREET:  One Liberty Place - 46th Floor
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  WORDPERFECT 5.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US91/01327
;      FILING DATE:  19910225
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  485,297
;      FILING DATE:  February 26, 1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jane Massey Licata
;      REGISTRATION NUMBER:  32,257
;      REFERENCE/DOCKET NUMBER:  ISIS-0085
```

```
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 568-3100
;      TELEFAX:  (215) 568-3439
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1557
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  unknown
PCT-US91-01327-12


alignment_scores:
            Quality:   86.50            Length:    165
              Ratio:  1.055              Gaps:      5
 Percent Similarity:  49.697   Percent Identity: 23.030

alignment_block:
 US-09-068-804-5 x PCT-US91-01327-12/rev   ..

 Align seg 1/1  to reverse of: PCT-US91-01327-12  from: 1  to: 1557

      94 ValAspArgLysArgGlnArgTrpProAlaAlaLeuLeuValArgLeuSe 110
         ::::::|||:::|||:::|||:::::::::::::|||:::      |||
    1485 CTCGGGCGGCAGCGACACCGACGACAGGACGGCCTTGTGCGTGCGCTG.. 1438

     110 rProLeuCysHisGlyAspAlaValIleValValValAlaValValGlyL 127
                                      |||:::   ::::::   :
    1437 ..............................GTAATTTATACACTGCTCC 1419

     127 ysGlyAlaAlaAlaLysLeuGlyAsnAlaLeuSerLysMetMetGlyGlu 143
         ::      |||::::::||||||
    1418 GTGAACGCGCGCCGAATCTTGGG........................... 1396

     144 ThrIleLysLysLeuValProAsnValLeuLysGlnLeuAlaGlnAsnGl 160
            |||   |||:::::::|||:::::::|||:::::::::   :::
    1395 ...ATTGCGAAGGTGGCGCCGGATGCCCTCCGGCACGTCATACGC..... 1354

     160 ySerLysLeuPheThrGlnGlyMetGlnArgIleThrSerGlyLeuGlyA 177
                                 |||   :::      |||||||||
    1353 .........................CAGGCCGTGGGTGTTGGTCTCGGCC 1329

     177 snValGlySerLysMetGlyLeuGlnThrAsnAlaLeuSerLysGluLeu 193
           |||:::          |||:::|||::::::|||::::::   |||
    1328 GAGTTGACAAAGAGGGCGGGGTGCAGAACGCGGCGATAGGCGAGGAGGGC 1279

     194 ValGlyAsnThr............LeuAsnLysValAlaLeuGlyMetGl 206
            |||::::::            :::   ||||||   ::::::   ::
    1278 CACGGCAAAGTCCGGCGAGAGCTGGTTGTTAAAGTACTGGTAGCCCGGGA 1229

     206 uValThrAsnThrAlaAlaGlnSerAlaGlyGlyValAlaGluGlyValP 223
         ::::   :::   :::||||||:::   |||   |||
    1228 CGCGGGTCACGGGGACGCCCAGGCTCGGGGCCACGTA.............. 1192

     223 heIleLysAsnAlaSerGluAlaLeuAlaAspPheMetLeuAlaArgPhe 239

    1191 ..............................................CAC 1189

     240 AlaMetAspGlnIleGlnGlnTrpLeuLysGlnSerValGluIle 254
         |||   :::|||:::||||||:::|||   |||:::||||||||
    1188 GCTAACCAGCAGCTCCAGCAGCGTCTGCCCCAGGGCGTAGAGATC 1144

seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-471-033-39

seq_documentation_block:
; Sequence 39, Application US/08471033
; Patent No. 5770696
;  GENERAL INFORMATION:
;    APPLICANT:  Warren, Gregory W
;    APPLICANT:  Koziel, Michael G
;    APPLICANT:  Mullins, Martha A
;    APPLICANT:  Nye, Gordon J
;    APPLICANT:  Carr, Brian
;    APPLICANT:  Desai, Nalini M
;    APPLICANT:  Kostichka, N. Kristy
;    APPLICANT:  Duck, Nicholas B
;    APPLICANT:  Estruch, Juan J
;    TITLE OF INVENTION:  No. 5770696el Pesticidal Proteins and Strains
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CIBA-GEIGY Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/471,033
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/314,594
;      FILING DATE:  09-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/218,018
;      FILING DATE:  23-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/037,057
;      FILING DATE:  25-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pace, Gary M.
;      REGISTRATION NUMBER:  P-40,403
;      REFERENCE/DOCKET NUMBER:  CGC 1695/CIP3/DIV7 - SQLv3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8582
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  39:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1241 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "Synthetic DNA"
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  9..1238
;      OTHER INFORMATION:  /note= "Maize optimized DNA
;      OTHER INFORMATION:  sequence encoding VIP2A(a) with the Bacillus secretion sig
;      OTHER INFORMATION:  removed as contained in pCIB5527"
US-08-471-033-39


alignment_scores:
            Quality:   82.50            Length:    259
              Ratio:  0.585              Gaps:      11
 Percent Similarity:  54.440   Percent Identity: 22.780

alignment_block:
 US-09-068-804-5 x US-08-471-033-39/rev   ..

 Align seg 1/1  to reverse of: US-08-471-033-39  from: 1  to: 1241

      14 GlyLysValLeuGlyAlaLeuLeuThrIleValSerValValAlaAlaVa 30
         |||   ||||||||:::   |||   ::::::      ||||||:::||
     890 GGTGATGTTCTCGGGGATGGGCTT...CTTGCCCAGGGCGTCGCTGATGT 844

      30 lPheThrGlyGlyAlaSerLeuAlaLeuAlaAlaValGlyLeuAlaValM 47
         |:::      |||:::   |||   :::|||||:::::|||:::::
     843 TCTTGATCTGGGCGTCCAGCTTCTCGTTGCCGCTGCCGCCCTGGTTGCGC 794
```

```
     47 etValAlaAspGluIleValLysAlaAlaThrGlyValSerPheIleGln 63
        |||:::   :::::::::   :::   :::||||||:::
    793 AGGTAGTTGTTGATCTCCTTGTAGTCCTGGCGGGCGTAGCC......GTC 750

     64 GlnAlaLeuAsnProIleMetGluHisValLeuLysProLeuMetGluLe 80
        |||:::|||:::      :::   :::||||||  ||||||:::   ::
    749 CAGGGCCTCGCGCTGGCTGTCGGTCAGGTCCTTGGCCCACTCCTCGTAGT 700

     80 u..............IleGlyLysAlaIleThrLysAlaLeuGlu.... 90
        :             :::|||   ::::::      :::||||||
    699 TCTTCATGCCCCAGCTGTGGGCCTCGGCGTTGATGTCGTTCTTGAAGTCT 650

     91 .................GlyLeuGlyValAspArgLysArgGlnArgTrp 101
                        :::|||:::::::::      :::
    649 AGACTCTTCTTCAGGGTGCCCTCGATCTGGAGGCACTCCAC......... 609

    102 ProAlaAlaLeuLeuValArgLeuSerProLeuCysHisGlyAspAlaVa 118
            |||||||||   :::|||:::   |||  |||   :::   ||
    608 ......GCCCTTCTTCACCACCTTGCTCACCTTGTCCACGTGCACCATGT 565

    118 lIleVal..........:.....ValValAlaValValGlyLys....... 127
        |   |||             |||:::|||||||||
    564 AGCCGTTGTCGATCAGCATCTTGTACTCGCTGTTGTTCAGGATCACGCCG 515

    128 ........GlyAlaAlaAlaLysLeuGlyAsnAlaLeuSerLysMetMet 141
                |||::::::|||   |||:::::::::   :::   :::
    514 GCCTTGGTGGGGGTGGTGCTGCCCTTGCCGCTGGGGACGGTCACCTTCAG 465

    142 GlyGluThrIleLysLysLeuValProAsnValLeuLysGlnLeuAlaGl 158
        ::::::::::::      |||:::::::::::|||
    464 GATCACGCGCTC......CTTGCTGCTCACCTGCTGGGC........... 432

    158 nAsnGlySerLysLeuPheThrGlnGlyMetGlnArgIleThrSerGlyL 175

    432 ..................................................... 432

    175 euGlyAsnValGlySerLysMetGlyLeuGlnThrAsnAlaLeuSerLys 191
          |||:::||||||   ::::::::::::::   ::::::      :::
    431 ..GGTCAGGTGGGTGTCCAGGTAGCTGTCGAACTTGATGTCGCGGTCCAG 384

    192 GluLeuVal..............GlyAsnThrLeuAsnLysValAlaLe 203
        ||||||:::               |||   :::::::::   ||||||||
    383 GAACTGCTCCTTGAACTGGGCCATGGCGTCGCTGTTGATGGTGTTGCCCT 334

    203 uGlyMetGluValThrAsnThrAlaAlaGlnSerAlaGlyGlyValAlaG 220
        ||||   ::::::::::::::::                     :::
    333 CGGTCAGGCTCTTGTTGAAGCCGAT.................GGTG 305

    220 luGlyValPheIleLysAsnAlaSerGluAlaLeuAlaAspPheMetLeu 236
          |||:::   :::      :::::::::|||:::|||:::      |||
    304 GTGGGCTCCACGTTCTTGTAGGTGATGATGCTGTTGCTCAGGTTGGTCTT 255

    237 AlaArgPheAlaMetAspGlnIleGln 245
        :::         :::|||   :::|||
    254 GTCGAACATCTTGTCGATCTCCTTCAG 228

seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-471-033-42

seq_documentation_block:
; Sequence 42, Application US/08471033
; Patent No. 5770696
;  GENERAL INFORMATION:
;    APPLICANT:  Warren, Gregory W
;    APPLICANT:  Koziel, Michael G
;    APPLICANT:  Mullins, Martha A
;    APPLICANT:  Nye, Gordon J
;    APPLICANT:  Carr, Brian
;    APPLICANT:  Desai, Nalini M
;    APPLICANT:  Kostichka, N. Kristy
;    APPLICANT:  Duck, Nicholas B
;    APPLICANT:  Estruch, Juan J
;    TITLE OF INVENTION:  No. 5770696el Pesticidal Proteins and Strains
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CIBA-GEIGY Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/471,033
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/314,594
;      FILING DATE:  09-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/218,018
;      FILING DATE:  23-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/037,057
;      FILING DATE:  25-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pace, Gary M.
;      REGISTRATION NUMBER:  P-40,403
;      REFERENCE/DOCKET NUMBER:  CGC 1695/CIP3/DIV7 - SQLv3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8582
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  42:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1241 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "Synthetic DNA"
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  9..1238
;      OTHER INFORMATION:  /note= "Maize optimized DNA
;      OTHER INFORMATION:  sequence encoding VIP2A(a) with the Bacillus secretion sig
;      OTHER INFORMATION:  removed and the eukaryotic secretion signal inserted as
;      OTHER INFORMATION:  contained in pCIB5528"
US-08-471-033-42


alignment_scores:
            Quality:   82.50            Length:    259
              Ratio:  0.585              Gaps:     11
 Percent Similarity:  54.440   Percent Identity: 22.780

alignment_block:
 US-09-068-804-5 x US-08-471-033-42/rev   ..

 Align seg 1/1  to reverse of: US-08-471-033-42  from: 1  to: 1241

     14 GlyLysValLeuGlyAlaLeuLeuThrIleValSerValValAlaAlaVa 30
        |||   |||||||||:::   |||   ::::::      ||||||:::||
    890 GGTGATGTTCTCGGGGATGGGCTT...CTTGCCCAGGGCGTCGCTGATGT 844

     30 lPheThrGlyGlyAlaSerLeuAlaLeuAlaAlaValGlyLeuAlaValM 47
        |:::      |||:::   |||   ::::||||||:::::::|||::::::
    843 TCTTGATCTGGGCGTCCAGCTTCTCGTTGCCGCTGCCGCCCTGGTTGCGC 794

     47 etValAlaAspGluIleValLysAlaAlaThrGlyValSerPheIleGln 63
          |||:::   :::::::::   :::   :::||||||:::
    793 AGGTAGTTGTTGATCTCCTTGTAGTCCTGGCGGGCGTAGCC......GTC 750
```

```
 64 GlnAlaLeuAsnProIleMetGluHisValLeuLysProLeuMetGluLe 80
    |||:::|||:::      :::   :::||||||   ||||||:::   ::
749 CAGGGCCTCGCGCTGGCTGTCGGTCAGGTCCTTGGCCCACTCCTCGTAGT 700

 80 u..............IleGlyLysAlaIleThrLysAlaLeuGlu.... 90
    :               :::|||   ::::::      :::||||||
699 TCTTCATGCCCCAGCTGTGGGCCTCGGCGTTGATGTCGTTCTTGAAGTCT 650

 91 ................GlyLeuGlyValAspArgLysArgGlnArgTrp 101
                    :::|||:::::::::      :::
649 AGACTCTTCTTCAGGGTGCCCTCGATCTGGAGGCACTCCAC......... 609

102 ProAlaAlaLeuLeuValArgLeuSerProLeuCysHisGlyAspAlaVa 118
        ||||||||   :::|||:::   |||   |||   :::   ||
608 ......GCCCTTCTTCACCACCTTGCTCACCTTGTCCACGTGCACCATGT 565

118 lIleVal..............ValValAlaValValGlyLys....... 127
    |   |||             |||:::||||||||||
564 AGCCGTTGTCGATCAGCATCTTGTACTCGCTGTTGTTCAGGATCACGCCG 515

128 ........GlyAlaAlaAlaLysLeuGlyAsnAlaLeuSerLysMetMet 141
          |||::::::|||   |||:::::::::   :::   :::
514 GCCTTGGTGGGGGTGGTGCTGCCCTTGCCGCTGGGGACGGTCACCTTCAG 465

142 GlyGluThrIleLysLysLeuValProAsnValLeuLysGlnLeuAlaGl 158
    ::::::::::::      |||:::::::::::::|||
464 GATCACGCGCTC......CTTGCTGCTCACCTGCTGGGC........... 432

158 nAsnGlySerLysLeuPheThrGlnGlyMetGlnArgIleThrSerGlyL 175

432 ............................................... 432

175 euGlyAsnValGlySerLysMetGlyLeuGlnThrAsnAlaLeuSerLys 191
      |||:::||||||   ::::::::::::::   ::::::      :::
431 ..GGTCAGGTGGGTGTCCAGGTAGCTGTCGAACTTGATGTCGCGGTCCAG 384

192 GluLeuVal..............GlyAsnThrLeuAsnLysValAlaLe 203
    ||||||:::               |||   :::::::::   ||||||||
383 GAACTGCTCCTTGAACTGGGCCATGGCGTCGCTGTTGATGGTGTTGCCCT 334

203 uGlyMetGluValThrAsnThrAlaAlaGlnSerAlaGlyGlyValAlaG 220
    ||||   ::::::::::::::::::                   :::
333 CGGTCAGGCTCTTGTTGAAGCCGAT.....................GGTG 305

220 luGlyValPheIleLysAsnAlaSerGluAlaLeuAlaAspPheMetLeu 236
      |||:::   :::      :::::::::|||:::|||:::      |||
304 GTGGGCTCCACGTTCTTGTAGGTGATGATGCTGTTGCTCAGGTTGGTCTT 255

237 AlaArgPheAlaMetAspGlnIleGln 245
    :::         :::|||   :::|||
254 GTCGAACATCTTGTCGATCTCCTTCAG 228
```

```
OM of: US-09-068-804-5 to: N_Geneseq_36:*   out_format : pfs

Date: Oct 5, 1999  8:01 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn2_1/USPTO_spool/US09068804/runat_05101999_145135_10341/app_query.fasta.1
 -DB=N_Geneseq_36 -QFMT=fastap -SUFFIX=rng -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=pam150
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=escore
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09068804
 -NCPU=6 -ICPU=3 -WAIT -THREADS=1

Search information block:
Query: US-09-068-804-5
Query length: 290
Database: N_Geneseq_36:*
Database sequences: 311585
Database length: 125096042
Search time (sec): 156.120000



score_list:
Sequence              Strd Orig    ZScore    EScore Len  ! Documentation  ..
N_Geneseq_36:T67035   + 1202.50 2517.54  1.3e-132  3441  ! Salmonella virulence factor Ss
N_Geneseq_36:T67028   + 1193.50 2512.65  2.4e-132  870  ! Salmonella secreted protein Ssp
N_Geneseq_36:T85939   +  557.00 1154.54  1.0e-56   1743  ! Shigella flexneri IpaB gene. S
N_Geneseq_36:Q31018   +  106.00  198.84  0.0018    1506  ! Thyroid N-acetyl-glucosamine r
N_Geneseq_36:T14618   +  101.50  186.63  0.0085    1947  ! DHFR/C. pneumoniae antigen fus
N_Geneseq_36:Q31005   +   91.50  181.41  0.0167    414  ! Thyroid N-acetyl-glucosamine re
N_Geneseq_36:T14612   +   95.50  176.85  0.0300    1464  ! Chlamydia pneumoniae polypepti
N_Geneseq_36:V22674   +  108.50  167.87  0.0948    50341  ! DNA sequence of a L5 shuttle
N_Geneseq_36:T51411   +  108.50  167.47  0.0997    52297  ! Mycobacteriophage L5 genome s
N_Geneseq_36:Q47357   +  108.50  167.47  0.0997    52298  ! L5 mycobacteriophage DNA. Myc
N_Geneseq_36:T64548   -   91.00  167.23  0.1028    1473  ! NADP-specific glutamate dehydr
N_Geneseq_36:T64547   -   91.00  167.00  0.1059    1506  ! NADP-specific glutamate dehydr
N_Geneseq_36:T64531   -   91.00  164.23  0.1511    1969  ! Glutamate dehydrogenase cDNA c
N_Geneseq_36:Q71390   -   95.00  163.92  0.1573    4614  ! Yeast 4.7 kB agglutination gen
N_Geneseq_36:T64543   -   91.00  163.59  0.1642    2096  ! NADP-specific glutamate dehydr
N_Geneseq_36:T64530   -   91.00  163.57  0.1645    2099  ! NADP-specific glutamate dehydr
N_Geneseq_36:T64542   -   91.00  163.39  0.1684    2137  ! NADP-specific glutamate dehydr
N_Geneseq_36:T64529   -   91.00  163.37  0.1688    2140  ! NADP-specific glutamate dehydr
N_Geneseq_36:Q64827   +   92.00  159.01  0.2954    4008  ! Leukotoxin AppIIIA gene. DNA e
N_Geneseq_36:X12986   +   96.00  156.78  0.3932    11307  ! Enterococcus faecalis genome
N_Geneseq_36:Q22246   -   81.50  153.25  0.6180    810  ! Sequence cDNA clone 26.j which
N_Geneseq_36:T73220   +   92.00  152.23  0.7045    7721  ! ApxIIIABCD gene. Preparations
N_Geneseq_36:T13948   -   82.50  150.96  0.8287    1241  ! Bacillus VIP2A(a) gene with eu
N_Geneseq_36:T13956   -   82.50  150.96  0.8287    1241  ! Maize-optimized VIP2A(a) gene
N_Geneseq_36:T74007   -   82.50  150.96  0.8287    1241  ! Maize optimised-B. cereus VIP2
N_Geneseq_36:T74008   -   82.50  150.96  0.8287    1241  ! Maize optimised-B. cereus VIP2
N_Geneseq_36:V16184   -   82.50  150.96  0.8287    1241  ! Maize optimised DNA sequence f
N_Geneseq_36:V16186   -   82.50  150.96  0.8287    1241  ! Maize optimised DNA sequence f
N_Geneseq_36:Q39018   -   87.50  150.93  0.8318    3474  ! Sequence of pure maize optimiz
N_Geneseq_36:T74012   -   87.50  150.93  0.8318    3474  ! Maize optimised-B. thuringiens
N_Geneseq_36:V16191   -   87.50  150.93  0.8318    3474  ! Maize optimised DNA encoding a
N_Geneseq_36:V73507   +   81.50  150.84  0.8421    1023  ! E. chrysanthemi HrpN-ECH DNA #
N_Geneseq_36:V69792   +   87.00  150.57  0.8710    3246  ! Candida albicans CaNik1 gene i
N_Geneseq_36:X21129   +   79.00  150.51  0.8783    632  ! Polynucleotide sequence from th
N_Geneseq_36:T13949   -   82.50  150.03  0.9338    1358  ! Bacillus VIP2A(a) gene with va
N_Geneseq_36:T74009   -   82.50  150.03  0.9338    1358  ! Maize optimised-B. cereus VIP2
N_Geneseq_36:V16188   -   82.50  150.03  0.9338    1358  ! Maize optimised DNA sequence f
N_Geneseq_36:Q65409   -   81.50  149.87  0.9530    1123  ! cDNA encoding Johnson grass pc
N_Geneseq_36:Q73596   -   81.50  149.86  0.9541    1124  ! Lolium perenne protein allerge
N_Geneseq_36:T13946   -   82.50  149.80  0.9622    1389  ! Maize-optimized VIP2A(a) gene.
N_Geneseq_36:T74002   -   82.50  149.80  0.9622    1389  ! Maize optimised-Bacillus VIP2A
N_Geneseq_36:V16175   -   82.50  149.80  0.9622    1389  ! Maize optimised DNA sequence f
N_Geneseq_36:T13953   -   82.50  149.72  0.9714    1399  ! Maize-optimized VIP2A(a) ge
N_Geneseq_36:T74000   -   82.50  149.72  0.9714    1399  ! Maize optimised-B. cereus V
N_Geneseq_36:V16173   -   82.50  149.72  0.9714    1399  ! Maize optimised DNA sequenc

seq_name: N_Geneseq_36:T67035

seq_documentation_block:
ID   T67035 standard; DNA; 3441 BP.
AC   T67035;
DT   04-AUG-1997  (first entry)
DE   Salmonella virulence factor SspBCDA DNA region.
KW   Salmonella secreted protein; SspB; SspC; SspD; SspA;
KW   bacterial-mediated endocytosis; diagnosis; therapy; vaccine;
KW   attenuation; virulence; ss.
OS   Salmonella typhimurium.
FH   Key             Location/Qualifiers
FT   cds             1. .870
FT                   /*tag= a
FT                   /product= SspB (incomplete)
FT   cds             898. .2127
FT                   /*tag= b
FT                   /product= SspC
FT   cds             2198. .3228
FT                   /*tag= c
FT                   /product= SspD
FT   cds             3247. .3441
FT                   /*tag= d
FT                   /product= SspA (incomplete)
PN   WO9718225-A1.
PD   22-MAY-1997.
PF   14-NOV-1996; U18504.
PR   14-NOV-1995; US-006733.
PA   (GEHO ) GEN HOSPITAL CORP.
PI   Miller SI;
DR   WPI; 97-289217/26.
DR   P-PSDB; W15268-71.
PT   New isolated Salmonella secreted proteins and related genes - used
PT   to develop products for the detection, treatment or prevention of
PT   Salmonella infections
PS   Claim 11; Fig 28; 95pp; English.
CC   A DNA molecule (T67035) comprises the SspBCDA locus (truncated at
CC   the 5' and 3' ends) including the partial sspB, sspC, sspD and
CC   partial sspA sequences (see also T67028-31) coding for Salmonella
CC   typhimurium secreted proteins SspB, SspC, SspD and SspA (W15268-
CC   70).  To identify the ssp genes, S. typhimurium mutants were
CC   investigated in invasion assays on HEp-2 epithelial cells.  2
CC   Ssp mutants with transposon insertions located between spaT and
CC   prgH (see also T67036) were identified.  One mutant lacked Ssp of
CC   87, 42 and 36 kDa.  The invasiveness of this mutant implicated the
CC   42 and 36 kDa Ssp in invasion.  DNA sequencing of this region
CC   identified the 4 ssp (partial) genes.  The ssp genes can be used in
CC   methods for diagnosing Salmonella infection and in methods of
CC   targetting an antigen to epithelial cells and of inducing a
CC   cytotoxic T cell immune response in a mammal.  Ssp gene mutation
CC   can be used to create attenuated Salmonella strains useful as live
CC   vaccines.
SQ   Sequence   3441 BP;    978 A;    761 C;    925 G;    777 T;


alignment_scores:
          Quality: 1202.50              Length:    292
            Ratio:  4.175                 Gaps:      4
 Percent Similarity:  98.630   Percent Identity: 98.288

alignment_block:
 US-09-068-804-5 x T67035   ..

 Align seg 1/1  to: T67035  from: 1  to: 3441

       1 ArgLysAlaGluGluThrAsnArgIleMetGlyCysIleGlyLysValLe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 CGCAAAGCCGAGGAAACGAACCGCATTATGGGATGTATCGGGAAAGTCCT 50
```

```
 17 uGlyAlaLeuLeuThrIleValSerValValAlaAlaValPheThrGlyG 34
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 51 CGGCGCGCTGCTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTG 100

 34 lyAlaSerLeuAlaLeuAlaAlaValGlyLeuAlaValMetValAlaAsp 50
    ||||||||||||||||||||||||||||||||||||||||||||||||||
101 GGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTGGCCGAT 150

 51 GluIleValLysAlaAlaThrGlyValSerPheIleGlnGlnAlaLeuAs 67
    ||||||||||||||||||||||||||||||||||||||||||||||||||
151 GAAATTGTGAAGGCGGCGACGGGAGTGTCGTTTATTCAGCAGGCGCTAAA 200

 67 nProIleMetGluHisValLeuLysProLeuMetGluLeuIleGlyLysA 84
    ||||||||||||||||||||||||||||||||||||||||||||||||||
201 CCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTGATTGGCAAGG 250

 84 laIleThrLysAlaLeuGluGlyLeuGlyValAsp.ArgLysArgGlnAr 100
    |||||||||||||||||||||||||||||||| ||||||||||||||||
251 CGATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAG 300

100 gTrpProAlaAlaLeuLeuValArgLeuSerProLeu.CysHisGlyAsp 116
    |||||||||||||||||||||||||||||||||| |||||||||
301 ATGGCCGGCAGCATTGTTGGTGCGATTGTCGCCGCTATTGCCATGGT... 347

117 Ala.ValIleValValValAlaValValGlyLysGlyAlaAlaAlaLysL 133
    ::: ||||||||||||||||||||||||||||||||||||||||||||||
348 GGCGGTCATTGTGGTGGTCGCAGTTGTCGGGAAAGGCGCGGCGGCGAAAC 397

133 euGlyAsnAlaLeuSerLysMetMetGlyGluThrIleLysLysLeuVal 149
    ||||||||||||||||||||||||||||||||||||||||||||||||||
398 TGGGTAACGCGCTGAGCAAAATGATGGGCGAAACGATTAAGAAGTTGGTG 447

150 ProAsnValLeuLysGlnLeuAlaGlnAsnGlySerLysLeuPheThrGl 166
    ||||||||||||||||||||||||||||||||||||||||||||||||||
448 CCTAACGTGCTGAAACAGTTGGCGCAAAACGGCAGCAAACTCTTTACCCA 497

166 nGlyMetGlnArgIleThrSerGlyLeuGlyAsnValGlySerLysMetG 183
    ||||||||||||||||||||||||||||||||||||||||||||||||||
498 GGGGATGCAACGTATTACTAGCGGTCTGGGTAATGTGGGTAGCAAGATGG 547

183 lyLeuGlnThrAsnAlaLeuSerLysGluLeuValGlyAsnThrLeuAsn 199
    ||||||||||||||||||||||||||||||||||||||||||||||||||
548 GCCTGCAAACGAATGCCTTAAGTAAAGAGCTGGTAGGTAATACCCTAAAT 597

200 LysValAlaLeuGlyMetGluValThrAsnThrAlaAlaGlnSerAlaGl 216
    ||||||||||||||||||||||||||||||||||||||||||||||||||
598 AAAGTGGCGTTGGGCATGGAAGTCACGAATACCGCAGCCCAGTCAGCCGG 647

216 yGlyValAlaGluGlyValPheIleLysAsnAlaSerGluAlaLeuAlaA 233
    ||||||||||||||||||||||||||||||||||||||||||||||||||
648 TGGTGTTGCCGAGGGCGTATTTATTAAAAATGCCAGCGAGGCGCTTGCTG 697

233 spPheMetLeuAlaArgPheAlaMetAspGlnIleGlnGlnTrpLeuLys 249
    ||||||||||||||||||||||||||||||||||||||||||||||||||
698 ATTTTATGCTCGCCCGTTTTGCCATGGATCAGATTCAGCAGTGGCTTAAA 747

250 GlnSerValGluIlePheGlyGluAsnGlnLysValThrAlaGluLeuGl 266
    ||||||||||||||||||||||||||||||||||||||||||||||||||
748 CAATCCGTAGAAATATTTGGTGAAAACCAGAAGGTAACGGCGGAACTGCA 797

266 nLysAlaMetSerSerAlaValGlnGlnAsnAlaAspAlaSerArgPheI 283
    ||||||||||||||||||||||||||||||||||||||||||||||||||
798 AAAAGCCATGTCTTCTGCGGTACAGCAAAATGCGGATGCTTCGCGTTTTA 847

283 leLeuArgGlnSerArgAla 289
    |||||||||||||||||||
848 TTCTGCGCCAGAGTCGCGCA 867
```

seq_name: N_Geneseq_36:T67028

seq_documentation_block:
ID T67028 standard; DNA; 870 BP.
AC T67028;
DT 04-AUG-1997 (first entry)
DE Salmonella secreted protein SspB DNA.
KW Salmonella secreted protein; SspB; bacterial-mediated endocytosis;
KW diagnosis; therapy; vaccine; attenuation; virulence; ss.
OS Salmonella typhimurium.
FH Key Location/Qualifiers
FT cds 1. .870
FT /*tag= a
FT /product= SspB (incomplete)
FT /note= "deduced amino acid residues from bases
FT 286-354 of the coding sequence to not
FT correspond to the SspB sequence given in
FT Fig 22"
PN WO9718225-A1.
PD 22-MAY-1997.
PF 14-NOV-1996; U18504.
PR 14-NOV-1995; US-006733.
PA (GEHO) GEN HOSPITAL CORP.
PI Miller SI;
DR WPI; 97-289217/26.
DR P-PSDB; W15268.
PT New isolated Salmonella secreted proteins and related genes - used
PT to develop products for the detection, treatment or prevention of
PT Salmonella infections
PS Claim 3; Fig 19; 95pp; English.
CC A DNA molecule (T67028) codes for (incomplete) Salmonella secreted
CC protein SspB (W15268), a virulence factor involved in bacterial
CC mediated endocytosis, which is able to translocate into human
CC epithelial cells. To identify Ssp genes, S. typhimurium mutants
CC were investigated in invasion assays on HEp-2 epithelial cells. 2
CC Ssp mutants with transposon insertions located between spaT and
CC prgH were identified. One mutant lacked Ssp of 87, 42 and 36 kDa.
CC The invasiveness of this mutant implicated the 42 and 36 kDa Ssp in
CC invasion. DNA sequencing of this region (see also T67035)
CC identified 2 complete and 2 partial open reading frames, designated
CC sspB, sspC (T67029), sspD (T67030) and sspA (T67031). These
CC nucleic acids can be used in methods for diagnosing Salmonella
CC infection and in methods of targetting an antigen to epithelial
CC cells and of inducing a cytotoxic T cell immune response in a
CC mammal. Ssp gene mutation can be used to create attenuated
CC Salmonella strains useful as live vaccines.
SQ Sequence 870 BP; 216 A; 178 C; 281 G; 195 T;

alignment_scores:
Quality: 1193.50 Length: 292
Ratio: 4.159 Gaps: 4
Percent Similarity: 98.288 Percent Identity: 97.945

alignment_block:
US-09-068-804-5 x T67028 ..

Align seg 1/1 to: T67028 from: 1 to: 870

```
  1 ArgLysAlaGluGluThrAsnArgIleMetGlyCysIleGlyLysValLe 17
    ||||||||||||||||||||||||||||||||||||||||||||||||||
  1 CGCAAAGCCGAGGAAACGAACCGCATTATGGGATGTATCGGGAAAGTCCT 50

 17 uGlyAlaLeuLeuThrIleValSerValValAlaAlaValPheThrGlyG 34
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 51 CGGCGCGCTGCTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTG 100

 34 lyAlaSerLeuAlaLeuAlaAlaValGlyLeuAlaValMetValAlaAsp 50
    ||||||||||||||||||||||||||||||||||||||||||||||||||
101 GGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTGGCCGAT 150

 51 GluIleValLysAlaAlaThrGlyValSerPheIleGlnGlnAlaLeuAs 67
    ||||||||||||||||||||||||||||||||||||||||||||||||||
151 GAAATTGTGAAGGCGGCGACGGGAGTGTCGTTTATTCAGCAGGCGCTAAA 200

 67 nProIleMetGluHisValLeuLysProLeuMetGluLeuIleGlyLysA 84
    ||||||||||||||||||||||||||||||||||||||||||||||||||
201 CCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTGATTGGCAAGG 250
```

```
 84 laIleThrLysAlaLeuGluGlyLeuGlyValAsp.ArgLysArgGlnAr 100
    |||||||||||||||||||||||||||||||||| |||||||||||||
251 CGATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAG 300

100 gTrpProAlaAlaLeuLeuValArgLeuSerProLeu.CysHisGlyAsp 116
    |||||||||||||||||||||||||||||||||| |||||||||
301 ATGGCCGGCAGCATTGTTGGIGCGATTGTCGCCGCTATTGCCATGGT... 347

117 Ala.ValIleValValValAlaValValGlyLysGlyAlaAlaAlaLysL 133
    ::: ||||||||||||||||||||||||||||||||||||||||||||
348 GGCGGTCATTGTGGTGGTCGCAGTTGTCGGGAAAGGCGCGGCGGCGAAAC 397

133 euGlyAsnAlaLeuSerLysMetMetGlyGluThrIleLysLysLeuVal 149
    ||||||||||||||||||||||||||||||||||||||||||||||||||
398 TGGGTAACGCGCTGAGCAAAATGATGGGCGAAACGATTAAGAAGTTGGTG 447

150 ProAsnValLeuLysGlnLeuAlaGlnAsnGlySerLysLeuPheThrGl 166
    ||||||||||||||||||||||||||||||||||||||||||||||||||
448 CCTAACGTGCTGAAACAGTTGGCGCAAAACGGCAGCAAACTCTTTACCCA 497

166 nGlyMetGlnArgIleThrSerGlyLeuGlyAsnValGlySerLysMetG 183
    ||||||||||||||||||||||||||||||||||||||||||||||||||
498 GGGGATGCAACGTATTACTAGCGGTCTGGGTAATGTGGGTAGCAAGATGG 547

183 lyLeuGlnThrAsnAlaLeuSerLysGluLeuValGlyAsnThrLeuAsn 199
    ||||||||||||||||||||||||||||||||||||||||||||||||||
548 GCCTGCAAACGAATGCCTTAAGTAAAGAGCTGGTAGGTAATACCCTAAAT 597

200 LysValAlaLeuGlyMetGluValThrAsnThrAlaAlaGlnSerAlaGl 216
    ||||||||||||||||||||||||||||||||||||||||||||||||||
598 AAAGTGGCGTTGGGCATGGAAGTCACGAATACCGCAGCCCAGTCAGCCGG 647

216 yGlyValAlaGluGlyValPheIleLysAsnAlaSerGluAlaLeuAlaA 233
    ||||||||||||||||||||||||||||||||||||||||||||||||||
648 TGGTGTTGCCGAGGGCGTATTTATTAAAAATGCCAGCGAGGCGCTTGCTG 697

233 spPheMetLeuAlaArgPheAlaMetAspGlnIleGlnGlnTrpLeuLys 249
    ||||||||||||||||||||||||||||||||||||||||||||||||||
698 ATTTTATGCTCGCCCGTTTTGCCATGGATCAGATTCAGCAGTGGCTTAAA 747

250 GlnSerValGluIlePheGlyGluAsnGlnLysValThrAlaGluLeuGl 266
    ||||||||||||||||||||||||||||||||||||||||||||||||||
748 CAATCCGTAGAAATATTTGGTGAAAACCAGAAGGTAACGGCGGAACTGCA 797

266 nLysAlaMetSerSerAlaValGlnGlnAsnAlaAspAlaSerArgPheI 283
    |||||||||||||||||||||||||||||||||||||||||||||||||
798 AAAAGCCATGTCTTCTGCGGTACAGCAAAATGCGGATGCTTCGGGTTTTA 847

283 leLeuArgGlnSerArgAla 289
    |||||||||||||||||||
848 TTCTGCGCCAGAGTCGCGCA 867
```

seq_name: N_Geneseq_36:T85939

seq_documentation_block:

ID T85939 standard; DNA; 1743 BP.
AC T85939;
DT 10-MAR-1998 (first entry)
DE Shigella flexneri IpaB gene.
KW IpaB; apoptosis; interleukin-1-beta converting enzyme; ICE; cacner;
KW autoimmune disease; inflammation; AIDS; degenerative disease;
KW Alzheimer's disease; myelodysplastic disorder; ischaemic injury;
KW toxin-induced liver disease; gene therapy; shigellosis; ss.
OS Shigella flexneri.
PN WO9726790-A1.
PD 31-JUL-1997.
PF 10-JAN-1997; U00399.
PR 25-JAN-1996; US-591079.
PA (UYNY) UNIV NEW YORK STATE.
PI Chen Y, Zychlinsky A;
DR WPI; 97-393276/36.
DR P-PSDB; W29893.
PT Shigella IpaB protein induces apoptosis by binding to
PT interleukin-1-beta-converting enzyme - useful for treating cancer,
PT auto-immunity, inflammation, etc. by eradication of unwanted cells
PS Claim 3; Page 67-69; 108pp; English.
CC This sequence comprises DNA encoding the IpaB protein (see W29893)
CC of Shigella flexneri. This protein induces apoptosis by binding to
CC interleukin-1-beta converting enzyme (ICE). A claimed method of
CC inducing apoptosis in a eukaryotic cell, thereby killing the cell,
CC comprises: (a) providing to a cell to be killed a DNA molecule (I)
CC in expressible form which encodes the Shigella IpaB protein or a
CC functional derivative; and (b) causing (I) to be expressed in the
CC cell, thereby producing the IpaB protein or derivative which
CC induces apoptosis and kills the cell. This approach is useful in
CC treating e.g. cancer, autoimmunity, inflammation and chronic viral
CC infections. Protein or peptide fragments, and the DNA encoding
CC them, can also be used to treat diseases that involve apoptotic
CC mechanisms in their pathogenesis, e.g. AIDS, degenerative diseases
CC such as Alzheimer's disease, myelodysplastic disorders, ischaemic
CC injury or toxin-induced liver disease. Methods of gene therapy
CC relying upon controlled expression of IpaB in a target cell are
CC disclosed. Also claimed are methods for: inhibiting the
CC interaction of an apoptosis-inducing protein or peptide with ICE;
CC detecting a compound capable of inhibiting the binding of IpaB to
CC ICE; screening a candidate protein for its ability to interact
CC with IpaB in a cell; and treatment of shigellosis. Homologues of
CC IpaB such as yopB of Yersinia entercolitica and Yersinia
CC pseudotuberculosis, and sipB of Salmonella typhimurium and
CC Salmonella typhi can also be used in the methods of the invention.
SQ Sequence 1743 BP; 632 A; 360 C; 302 G; 449 T;

alignment_scores:

```
            Quality:  557.00              Length:    289
              Ratio:  2.411                 Gaps:      1
 Percent Similarity:  79.931   Percent Identity: 46.021
```

alignment_block:

US-09-068-804-5 x T85939 ..

Align seg 1/1 to: T85939 from: 1 to: 1743

```
   1 ArgLysAlaGluGluThrAsnArgIleMetGlyCysIleGlyLysValLe 17
     |||||||||||||||   ||||||:::|||||||||:::|||||:::||
 892 CGTAAAGCAGAAGAACTCAACAGAGTAATGGGTTGTGTTGGGAAAATACT 941

  17 uGlyAlaLeuLeuThrIleValSerValValAlaAlaValPheThrGlyG 34
     ||||||||||||||||||||||||||||||||||:::|||:::||||
 942 TGGGGCACTTTTAACTATCGTTAGTGTTGTTGCAGCAGCTTTTTCTGGAG 991

  34 lyAlaSerLeuAlaLeuAlaAlaValGlyLeuAlaValMetValAlaAsp 50
     ||||||||||||||||||||||||||||||||:::|||||:::||||
 992 GAGCCTCTCTAGCACTGGCAGCTGTTGGTTTAGCTCTTATGGTTACGGAT 1041

  51 GluIleValLysAlaAlaThrGlyValSerPheIleGlnGlnAlaLeuAs 67
     :::|||||||:::|||||||||||||   ||||||::::::||||||||||
1042 GCTATAGTACAAGCAGCGACCGGCAATTCCTTCATGGAACAAGCCCTGAA 1091

  67 nProIleMetGluHisValLeuLysProLeuMetGluLeuIleGlyLysA 84
     |||||||||      |||:::  ||||||:::   |||:::::::  |
1092 TCCGATCATGAAAGCAGTCATTGAACCCTTAATCAAACTCCTTTCAGATG 1141

  84 laIleThrLysAlaLeuGluGlyLeuGlyValAspArgLysArgGlnArg 100
     ||:::||||||   ||||||||||||||||||   |||:::   :::
1142 CATTTACAAAAATGCTCGAAGGCTTGGGCGTCGACTCGAAAAAAGCCAAA 1191

 101 TrpProAlaAlaLeuLeuValArgLeuSerProLeuCysHisGlyAspAl 117
          ::::::::::|||      ::::::                   ||
1192 ATGATTGGCTCTATTCTGGGGGCAATCGCAGGCGCTCTTGTCCTAGTTGC 1241

 117 aValIleValValValAlaValValGlyLysGlyAlaAlaAlaLysLeuG 134
     |:::::|||:::||||||:::|||||||||   ||||||||||||||:
1242 AGCAGTCGTTCTCGTAGCCACTGTTGGTAAACAGGCAGCAGCAAAACTTG 1291
```

```
     134 lyAsnAlaLeuSerLysMetMetGlyGluThrIleLysLysLeuValPro 150
         ::::::::::::::::|||:::::::|||   |||::::::   |||:::|||
    1292 CAGAAAATATTGGCAAAATAATAGGTAAAACCCTCACAGACCTTATACCA 1341

     151 AsnValLeuLysGlnLeuAlaGlnAsnGlySerLysLeuPheThrGlnGl 167
         :::   |||||:::::::::   :::   :::   |||:::|||:::::
    1342 AAGTTTCTCAAGAATTTTTCTTCTCAACTGGACGATTTAATCACTAATGC 1391

     167 yMetGlnArgIleThrSerGlyLeuGlyAsnValGlySerLysMetGlyL 184
         ::::   |||::::::      |||||:::::::|||:::
    1392 TGTTGCCAGATTAAATAAATTTCTTGGTGCAGCGGGTGAT.......... 1431

     184 euGlnThrAsnAlaLeuSerLysGluLeuValGlyAsnThrLeuAsnLys 200
                 :::::::::||||||:::::::::::::::   ||||||:::
    1432 ........GAAGTAATATCCAAACAAATTATTTCCACCCATTTAAACCAA 1473

     201 ValAlaLeuGlyMetGluValThrAsnThrAlaAlaGlnSerAlaGlyGl 217
         :::::|||      |||   :::|||:::|||:::|||::::::|||::
    1474 GCAGTTTTATTAGGAGAAAGTGTTAACTCTGCCACACAAGCGGGAGGAAG 1523

     217 yValAlaGluGlyValPheIleLysAsnAlaSerGluAlaLeuAlaAspP 234
         :||||||   :::||||||   ::::::||||||   :::|||||||||:
    1524 TGTCGCTTCTGCTGTTTTCCAGAACAGCGCGTCGACAAATCTAGCAGACC 1573

     234 heMetLeuAlaArgPheAlaMetAspGlnIleGlnGlnTrpLeuLysGln 250
         ::   |||::::::::::   :::::::|||:::   :::   :::   :::
    1574 TGACATTATCGAAATATCAAGTTGAACAACTGTCAAAATATATCAGTGAA 1623

     251 SerValGluIlePheGlyGluAsnGlnLysValThrAlaGluLeuGlnLy 267
         :::::::|||   ||||||:::   |||   |||:::|||:::|||
    1624 GCAATAGAAAAATTCGGCCAATTGCAGGAAGTAATTGCAGATCTATTAGC 1673

     267 sAlaMetSerSerAlaValGlnGlnAsnAlaAspAlaSerArgPheIleL 284
          :::||||||::::::      :::   :::|||:::::::::   ||||
    1674 CTCAATGTCCAACTCTCAGGCTAATAGAACTGATGTTGCAAAAGCAATTT 1723

     284 euArgGlnSerArgAla 289
         ||:::|||:::   |||
    1724 TGCAACAAACTACTGCT 1740
```

```
seq_name: N_Geneseq_36:Q31018

seq_documentation_block:
ID   Q31018 standard; cDNA; 1506 BP.
AC   Q31018;
DT   06-APR-1993  (first entry)
DE   Thyroid N-acetyl-glucosamine receptor clone TGR-CLH.
KW   GlcNAc receptor; carbohydrate recognition domain;
KW   thyroid adenocarcinoma; lysosome; endosome;
KW   Hashimoto's disease; Basedow's disease; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             16. .1011
FT                   /*tag= a
FT                   /product= GlcNAc_receptor
FT                   /note= "deduced amino acid sequence in patent
FT                   specification decodes sequence from
FT                   nucleotide 1 (vector-derived sequence)
FT                   and continues decoding beyond any nonsense
FT                   codons to nucleotide 1506"
FT   signal_peptide  16. .60
FT                   /*tag= b
FT                   /note= "putative"
FT   misc_feature    199. .594
FT                   /*tag= c
FT                   /label= TGR-CL1bis
FT                   /note= "see Q31005"
FT   misc_feature    463. .837
FT                   /*tag= d
FT                   /label= TGR-CL5bis
FT                   /note= "see Q31007"
PN   WO9219733-A.
PD   12-NOV-1992.
PF   30-APR-1992; F00396.
PR   03-MAY-1991; FR-005478.
PA   (CNRS ) CENT NAT RECH SCI.
PI   Blanck O, Courageot J, Miquelis R, Thibault V;
DR   WPI; 92-398860/48.
DR   P-PSDB; R32762.
PT   New thyroid N-acetyl-glucosamine receptor proteins - and
PT   fragments, antibodies and encoding nucleic acids, for therapeutic
PT   or diagnostic use
PS   Claim 7; Fig 7; 75pp; French.
CC   Denatured GlcNAc receptor was purified from porcine thyroids and
CC   used to raise polyclonal antibodies in rabbits. The antibody "TGRD-
CC   Ab" was used to screen a Western blot derived from a cDNA bank of
CC   normal human thyroid cDNA in lambda gt11. After 3 rounds of
CC   immunoscreening, 13 clones were confirmed as positive. Their EcoRI
CC   insert was sub-cloned in pUC18 and clones having ORFs of over 123bp
CC   (i.e clones TGR-CL1bis, -CL4, -CL5bis, -CL6, -CL7, -CL10 and -CL11)
CC   were sequenced. Clone TGR-CL1bis was used with TGR-CL5bis to screen
CC   the human cDNA library for longer clones; the clone TGR-CLH (Q31018)
CC   was isolated. The preferred DNA sequences of the invention which
CC   encode GlcNAc receptor protein fragments are useful as primers,
CC   e.g. to detect congenital thyroid disorders such as goitre and
CC   juvenile hypothyroidism. See Q31005-Q31018.
SQ   Sequence   1506 BP;    328 A;    325 C;    520 G;    333 T;

alignment_scores:
            Quality:  106.00              Length:    248
              Ratio:  0.930                 Gaps:     11
 Percent Similarity:  45.968   Percent Identity: 20.565

alignment_block:
 US-09-068-804-5 x Q31018   ..

 Align seg 1/1  to: Q31018  from: 1  to: 1506

      10 MetGlyCysIleGlyLysValLeuGlyAlaLeuLeuThrIleValSerVa 26
         ||||||||||||
      16 ATGGGTTGCATA..................................... 27

      26 lValAlaAlaValPheThrGlyGlyAlaSerLeuAlaLeuAlaAlaValG 43
                           |||   |||:::::::::::::::::|||:::|
      28 ...................GGTCCCGCAGGAATGGGAATGGAAGCCATAG 58

      43 lyLeuAlaVal..............MetValAla.............. 49
         ||:::::::::              |||   :::
      59 GATTTGGAATAAATAAAATGGGAGGAATGGAGGGCCCTTTGGTGGTGGT 108

      50 ...AspGluIleValLysAlaAlaThrGlyValSerPheIleGlnGlnAl 65
            :::::::::   :::   ::::::|||::::::      :::
     109 ATGGAAAACATGGGTCGATTTGGATCTGGGATGAACATGGGCAGG..... 153

      65 aLeuAsnProIleMetGluHisValLeuLys.................. 75
          :::|||   |||:::   ::::::||||||
     154 .ATAAATGAAATCCTAAGTAATGCACTGAAGAGAGGAGAGATCATTGCAA 202

      76 ..............................ProLeuMetGluLeuIle 81
                                       |||   :::|||   :::
     203 AGCAGGGAGGAGGTGGAGGTGGAAGCAGCGTCCCTGGGATCGAGAGGATG 252

      82 GlyLysAlaIleThrLysAlaLeuGluGlyLeuGlyValAspArgLysAr 98
         |||   :::|||   :::   |||   |||   |||::::::|||
     253 GGTCCTGGCATTGACCGC...CTCGGGGGTGCCGGCATGGAGCGC..... 294

      98 gGlnArgTrpProAlaAlaLeuLeuValArgLeuSerProLeuCysHisG 115

     294 ..................................................... 294

     115 lyAspAlaValIleValValValAlaValValGlyLysGly......... 128
                      :::   |||   :::|||   |||
     295 ..............ATCGGCGCGGGCCTGGGCCACGGCATGGATCGC 327

     129 .........AlaAlaAlaLysLeuGlyAsnAlaLeuSerLysMet..... 140
```

```
                 :::   ::::::::::|||   ::::::::::::|||
     328 GTGGGCTCCGAGATCGAGCGCATGGGCCTGGTCATGGACCGCATGGGCTC 377

     141 ..........MetGlyGluThrIleLysLysLeuValProAsnValLeuL 154
                  ||||||      |||   ::::::   |||      |||
     378 CGTGGAGCGCATGGGCTCCGGCATTGAGCGCATGGGCCCGCTGGGCCTCG 427

     154 ysGlnLeuAlaGlnAsnGlySerLysLeuPheThrGlnGlyMetGlnArg 170
         :::::::|||   :::      ::::::      |||   |||:::|||
     428 ACCACATGGCCTCCAGCATTGAGCGCATG...GGCCAGACCATGGAGCGC 474

     171 IleThrSerGlyLeuGlyAsnValGlySerLysMet......GlyLeuGl 185
         |||   ||||||:::      :::|||:::   |||      ||||||::
     475 ATTGGCTCTGGCGTGGAGCGCATGGGTGCCGGCATGGGCTTCGGCCTTGA 524

     185 nThrAsnAlaLeuSerLysGluLeuValGlyAsnThrLeuAsnLysValA 202
         :      |||   :::   :::   ||||||:::|||:::::::::::::
     525 GCGCATGGCCGCTCCCATCGACCGTGTGGGCCAGACCATTGAGCGCATGG 574

     202 laLeuGlyMetGluValThrAsnThrAlaAlaGlnSerAlaGly 216
         ::   |||:::|||      ::::::|||   :::      |||
     575 GCTCTGGCGTGGAGCGCATGGGCCCTGCCATCGAGCGCATGGGC 618

seq_name: N_Geneseq_36:T14618

seq_documentation_block:
ID   T14618 standard; DNA; 1947 BP.
AC   T14618;
DT   11-NOV-1996  (first entry)
DE   DHFR/C. pneumoniae antigen fusion protein (polypeptide B) DNA.
KW   Polypeptide antigen; polypeptide B; strain YK41; plasmid; probe;
KW   pCPN533T; primer; assay; detection; antibody; diagnosis;
KW   infection; fusion protein; dihydrofolate reductase; DHFR;
KW   Chlamydia pneumoniae; ds.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   cds             1. .1947
FT                   /*tag=  a
FT                   /note= "STOP codon absent"
PN   WO9609320-A1.
PD   28-MAR-1996.
PF   20-SEP-1995; J01896.
PR   20-SEP-1994; JP-224711.
PR   28-APR-1995; JP-106011.
PR   28-APR-1995; JP-106006.
PR   28-APR-1995; JP-106010.
PR   28-APR-1995; JP-106008.
PR   28-APR-1995; JP-106009.
PA   (HITB ) HITACHI CHEM CO LTD.
PI   Izutsu H,  Matsumoto A,  Obara K;
DR   WPI; 96-188399/19.
DR   P-PSDB; R94584.
PT   Recombinant Chlamydia pneumoniae antigen and antibodies to it -
PT   used for detection and assay of C. pneumoniae e.g. in clinical
PT   diagnosis
PS   Claim 25; Pages 97-103; 128pp; Japanese.
CC   The present sequence encodes the dihydrofolate reductase (DHFR)/
CC   C. pneumoniae polypeptide antigen fusion protein, polypeptide B.
CC   C. pneumoniae strain YK41 was cultured and genomic DNA extracted
CC   to prep. a lambda gt11 DNA library. The library was then screened
CC   with an anti-YK41 monoclonal antibody (MAb), which was prepd. by
CC   fusing spleen cells from a mouse infected with YK41 with myeloma
CC   P3/NS1/1-Ag4-1 to produce a MAb expressing hydridoma. The DNA
CC   obtd. was then fused with DHFR DNA and the expression vector
CC   pADA431 to give pCPN533T. The plasmid was used to transform an
CC   E. coli host, which was cultured to give the antigenic polypeptide
CC   fusion protein, polypeptide B. Polypeptide B and primers and
CC   probes derived from its DNA can be used in assays for the
CC   detection of antigenic polypeptide antibodies and DNA,
CC   respectively, useful in the diagnosis of C. pneumoniae infection.
SQ   Sequence   1947 BP;    591 A;    452 C;    483 G;    421 T;

alignment_scores:
            Quality:  101.50            Length:    227
              Ratio:  0.839              Gaps:      6
 Percent Similarity:  53.304   Percent Identity: 23.789

alignment_block:
 US-09-068-804-5 x T14618   ..

 Align seg 1/1  to: T14618  from: 1  to: 1947

       1 ArgLysAlaGluGluThrAsnArgIleMetGlyCysIleGlyLysValLe 17
         :::|||:::   :::   :::   :::|||:::   :::::::::|||::
    1204 CAGAAGTCTAAAGATCTCGAAGGAACAATGGATACTGTCAATACTGTGAT 1253

      17 uGlyAlaLeu.........LeuThrIleValSerValValAlaAlaValP 31
         :   |||:::          :::|||::::::|||:::|||||||||:::|
    1254 GATCGCGGTTTCTGTTGCCATTACAGTTATTTCTATTGTTGCTGCTATTT 1303

      31 heThrGlyGlyAlaSerLeu.............AlaLeuAlaAlaValGly 43
         |||||   ||||||:::|||             |||   ||||||||||||
    1304 TTACATGCGGAGCTGGACTCGCTGGACTCGCTGCGGGAGCTGCTGTAGGT 1353

      44 LeuAlaValMetValAlaAspGluIleValLysAlaAlaThrGlyValSe 60
         |||:::      :::::::::   :::   |||||||||   |||::
    1354 GCAGCGGCAGCTGGAGGTGCAGCAGGAGCTGCTGCCGCAACCACGGTAGC 1403

      60 rPheIleGlnGlnAlaLeuAsnProIleMetGluHisValLeuLysProL 77
         :   :::|||   :::::::::::::::::::::::   |||   :::::::
    1404 A...ACACAAATTACAGTTCAAGCTGTTGTCCAAGCGGTGAAACAAGCTG 1450

      77 euMetGluLeuIleGlyLysAlaIleThrLysAlaLeuGluGlyLeuGly 93
         :::::      :::   :::|||||||||   |||:::
    1451 TTATCACAGCTGTCAGACAAGCGATCACCGCGGCTATAAAA..... 1491

      94 ValAspArgLysArgGlnArgTrpProAlaAlaLeuLeuValArgLeuSe 110

    1491 ..................................................... 1491

     110 rProLeuCysHisGlyAspAlaValIleValValValAlaValValGlyL 127

    1491 ..................................................... 1491

     127 ysGlyAlaAlaAlaLysLeuGlyAsnAlaLeuSerLysMetMetGlyGlu 143
         |||||:::|||   |||                   :::   :::
    1492 .....GCGGCTGTCAAATCTGGA.................ATAAAAGCA 1518

     144 ThrIleLysLysLeuValProAsnValLeuLysGlnLeuAlaGlnAsnGl 160
         ||||||:::||||||   ::::::   |||   ::::::::::::
    1519 TTTATCAAAACTTTAGTCAAAGCGATTGCCAAAGCCATTTCTAAAGGAAT 1568

     160 ySerLysLeuPheThrGlnGlyMetGlnArgIleThrSerGlyLeuGlyA 177
          ||||||:::|||:::::::|||   |||   |||:::   ::::::   :
    1569 CTCTAAGGTTTTCGCTAAGGGAACTCAAATGATTGCGAAGAACTTCCCCA 1618

     177 snValGlySerLysMetGlyLeuGlnThrAsnAlaLeuSerLysGluLeu 193
         ::                                  |||||||||   :::
    1619 AG..............................CTCTCGAAAGTCATC 1635

     194 ValGlyAsnThrLeuAsnLysValAlaLeuGlyMetGluValThrAsnTh 210
         :::   |||   :::   |||:::::::|||:::   |||:::   ::
    1636 TCGTCTCTTACCAGTAAATGGGTCACGGTTGGGGTTGGGGTTGTAGTTGC 1685

     210 rAlaAlaGlnSerAlaGlyGlyValAlaGlu 220
         :|||:::      :::   |||:::   :::
    1686 GGCGCCTGCTCTCGGTAAAGGGATTATGCAA 1716

seq_name: N_Geneseq_36:Q31005

seq_documentation_block:
ID   Q31005 standard; cDNA; 414 BP.
AC   Q31005;
DT   06-APR-1993  (first entry)
DE   Thyroid N-acetyl-glucosamine receptor clone TGR-CL1bis.
```

```
KW   GlcNAc receptor; carbohydrate recognition domain;
KW   thyroid adenocarcinoma; lysosome; endosome;
KW   Hashimoto's disease; Basedow's disease; ss.
OS   Homo sapiens.
PN   WO9219733-A.
PD   12-NOV-1992.
PF   30-APR-1992; F00396.
PR   03-MAY-1991; FR-005478.
PA   (CNRS ) CENT NAT RECH SCI.
PI   Blanck O, Courageot J, Miquelis R, Thibault V;
DR   WPI; 92-398860/48.
DR   P-PSDB; R28981.
PT   New thyroid N-acetyl-glucosamine receptor proteins - and
PT   fragments, antibodies and encoding nucleic acids, for therapeutic
PT   or diagnostic use
PS   Claim 7; Fig 2a; 75pp; French.
CC   Denatured GlcNAc receptor was purified from porcine thyroids and
CC   used to raise polyclonal antibodies in rabbits. The antibody "TGRD-
CC   Ab" was used to screen a Western blot derived from a cDNA bank of
CC   normal human thyroid cDNA in lambda gt11. After 3 rounds of
CC   immunoscreening, 13 clones were confirmed as positive. Their EcoRI
CC   insert was sub-cloned in pUC18 and clones having ORFs of over 123bp
CC   (i.e clones TGR-CL1bis, -CL4, -CL5bis, -CL6, -CL7, -CL10 and -CL11)
CC   were sequenced. Clone TGR-CL1bis was used with TGR-CL5bis to screen
CC   the human cDNA library for longer clones; the clone TGR-CLH (Q31018)
CC   was isolated. The preferred DNA sequences of the invention which
CC   encode GlcNAc receptor protein fragments are useful as primers,
CC   e.g. to detect congenital thyroid disorders such as goitre and
CC   juvenile hypothyroidism. See Q31006-Q31017.
SQ   Sequence   414 BP;    64 A;    114 C;    166 G;    70 T;


alignment_scores:
            Quality:   91.50              Length:    138
              Ratio:  1.204                Gaps:      6
 Percent Similarity:  55.072   Percent Identity: 25.362

alignment_block:
 US-09-068-804-5 x Q31005   ..

 Align seg 1/1  to: Q31005  from: 1  to: 414

      76 ProLeuMetGluLeuIleGlyLysAlaIleThrLysAlaLeuGluGlyLe 92
         |||   :::|||   :::|||   :::|||   :::   |||   |||
      46 CCTGGGATCGAGAGGATGGGTCCTGGCATTGACCGC...CTCGGGGGTGC 92

      92 uGlyValAspArgLysArgGlnArgTrpProAlaAlaLeuLeuValArgL 109
         |||::::::|||:::
      93 CGGCATGGAGCGCATGGGC..............................  111

     109 euSerProLeuCysHisGlyAspAlaValIleValValValAlaAlaVal 125
           :::   |||   ||||||   :::   ::::::
     112 ..GCGGGCCTGGGCCACGGCATGGATCGCGTGGCT..............  144

     126 GlyLysGlyAlaAlaAlaLysLeuGlyAsnAlaLeuSerLysMet.....  140
             ::::::   ::::::::::|||   ::::::::::::|||
     145 ......TCCGAGATCGAGCGCATGGGCCTGGTCATGGACCGCATGGGCTC 188

     141 ..........MetGlyGluThrIleLysLysLeuValProAsnValLeuL 154
                   ||||||      |||   ::::::   |||      |||
     189 CGTGGAGCGCATGGGCTCCGGCATTGAGCGCATGGGCCCGCTGGGCCTCG 238

     154 ysGlnLeuAlaGlnAsnGlySerLysLeuPheThrGlnGlyMetGlnArg 170
           ::::::|||   :::      ::::::      |||   |||:::|||
     239 ACCACATGGCCTCCAGCATTGAGCGCATG...GGCCAGACCATGGAGCGC 285

     171 IleThrSerGlyLeuGlyAsnValGlySerLysMet......GlyLeuGl 185
         |||   ||||||:::      :::|||:::   |||      ||||||::
     286 ATTGGCTCTGGCGTGGAGCGCATGGGTGCCGGCATGGGCTTCGGCCTTGA 335

     185 nThrAsnAlaLeuSerLysGluLeuValGlyAsnThrLeuAsnLysValA 202
         :      |||   :::   :::   ||||||:::|||:::::::::::::
     336 GCGCATGGCCGCTCCCATCGACCGTGTGGGCCAGACCATTGAGCGCATGG 385

     202 laLeuGlyMetGlu 206
         ::   |||:::|||
     386 GCTCTGGCGTGGAG 399

seq_name: N_Geneseq_36:T14612

seq_documentation_block:
ID   T14612 standard; DNA; 1464 BP.
AC   T14612;
DT   07-NOV-1996  (first entry)
DE   Chlamydia pneumoniae polypeptide antigen (polypeptide A) DNA.
KW   Polypeptide antigen; polypeptide A; strain YK41; plasmid; probe;
KW   pCPN533alpha; primer; assay; detection; antibody; diagnosis;
KW   infection; ds.
OS   Chlamydia pneumoniae.
FH   Key             Location/Qualifiers
FT   cds             1. .1464
FT                   /*tag=  a
FT                   /note= "STOP codon absent"
FT   misc_feature    1. .777
FT                   /*tag=  b
FT                   /note= "claimed 5'-end fragment"
PN   WO9609320-A1.
PD   28-MAR-1996.
PF   20-SEP-1995; J01896.
PR   20-SEP-1994; JP-224711.
PR   28-APR-1995; JP-106011.
PR   28-APR-1995; JP-106006.
PR   28-APR-1995; JP-106010.
PR   28-APR-1995; JP-106008.
PR   28-APR-1995; JP-106009.
PA   (HITB ) HITACHI CHEM CO LTD.
PI   Izutsu H,  Matsumoto A,  Obara K;
DR   WPI; 96-188399/19.
DR   P-PSDB; R94579.
PT   Recombinant Chlamydia pneumoniae antigen and antibodies to it  -
PT   used for detection and assay of C. pneumoniae e.g. in clinical
PT   diagnosis
PS   Claim 9; Pages 60-64; 128pp; Japanese.
CC   The present sequence encodes the C. pneumoniae polypeptide antigen,
CC   polypeptide A. C. pneumoniae strain YK41 was cultured and genomic
CC   DNA extracted to prep. a lambda gt11 DNA library. The library was
CC   then screened with an anti-YK41 monoclonal antibody (MAb), which
CC   was prepd. by fusing spleen cells from a mouse infected with YK41
CC   with myeloma P3/NS1/1-Ag4-1 to produce a MAb expressing hydridoma.
CC   The DNA obtd. was then fused with the expression vector pADA431 to
CC   give pCPN533alpha. The plasmid was used to transform an E. coli
CC   host, which was cultured to give the antigenic polypeptide,
CC   polypeptide A. Polypeptide A and primers and probes derived from
CC   its DNA can be used in assays for the detection of polypeptide A
CC   antibodies and DNA, respectively, useful in the diagnosis of
CC   C. pneumoniae infection.
SQ   Sequence   1464 BP;    470 A;    333 C;    346 G;    315 T;


alignment_scores:
            Quality:   95.50              Length:    227
              Ratio:  0.796                Gaps:      6
 Percent Similarity:  52.863   Percent Identity: 23.348

alignment_block:
 US-09-068-804-5 x T14612   ..

 Align seg 1/1  to: T14612  from: 1  to: 1464

       1 ArgLysAlaGluGluThrAsnArgIleMetGlyCysIleGlyLysValLe 17
         :::|||:::   :::   :::   :::|||:::   ::::::::::|||::
     721 CAGAAGTCTAAAGATCTCGAAGGAACAATGGATACTGTCAATACTGTGAT 770

      17 uGlyAlaLeu.........LeuThrIleValSerValValAlaAlaValP 31
         :   |||:::          :::|||::::::|||:::|||||||||:::|
     771 GATCGCGGTTTCTGTTGCCATTACAGTTATTTCTATTGTTGCTGCTATTT 820
```

```
     31 heThrGlyGlyAlaSerLeu.....AlaLeuAlaAlaValGly 43
     821 TTACATGCGGAGCTGGACTCGCTGGACTCGCTGCGCGAGCTGCTGTACGT 870

     44 LeuAlaValMetValAlaAspGluIleValLysAlaAlaThrGlyValSe 60
     871 GCAGCGGCAGCTGGAGGTGCAGCAGGAGCTGCTGCCGCAACCACGGTAGC 920

     60 rPheIleGlnGlnAlaLeuAsnProIleMetGluHisValLeuLysProL 77
     921 A...ACACAAATTACAGTTCAAGCTGTTGTCCAAGCGGTGAAACAAGCTG 967

     77 euMetGluLeuIleGlyLysAlaIleThrLysAlaLeuGluGlyLeuGly 93
     968 TTATCACAGCTGTCAGACAAGCGATCACCGCGGCTATAAAA..... 1008

     94 ValAspArgLysArgGlnArgTrpProAlaAlaLeuLeuValArgLeuSe 110
    1008 ..... 1008

    110 rProLeuCysHisGlyAspAlaValIleValValValAlaValValGlyL 127
    1008 ..... 1008

    127 ysGlyAlaAlaAlaLysLeuGlyAsnAlaLeuSerLysMetMetGlyGlu 143
    1009 .....GCGGCTGTCAAATCTGGA.....ATAAAAGCA 1035

    144 ThrIleLysLysLeuValProAsnValLeuLysGlnLeuAlaGlnAsnGl 160
    1036 TTTATCAAAACTTTAGTCAAAGCGATTGCCAAAGCCATTTCTAAAGGAAT 1085

    160 ySerLysLeuPheThrGlnGlyMetGlnArgIleThrSerGlyLeuGlyA 177
    1086 CTCTAAGGTTTCGCTAAGGGAACTCAAATGATTGCGAAGAACTTCCCCA 1135

    177 snValGlySerLysMetGlyLeuGlnThrAsnAlaLeuSerLysGluLeu 193
    1136 AG.....CTCTCGAAAGTCATC 1152

    194 ValGlyAsnThrLeuAsnLysValAlaLeuGlyMetGluValThrAsnTh 210
    1153 TCGTCTCTTACCAGTAAATGGGTCACGGTTGGGGTTGGGGTTGTAGTTGC 1202

    210 rAlaAlaGlnSerAlaGlyGlyValAlaGlu 220
    1203 GGCGCCTGCTCTCGGTAAAGGGATTATGCAA 1233
```

seq_name: N_Geneseq_36:V22674

seq_documentation_block:

```
ID   V22674 standard; DNA; 50341 BP.
AC   V22674;
DT   15-JUL-1998  (first entry)
DE   DNA sequence of a L5 shuttle phasmid designated phAE41.
KW   L5 shuttle phasmid; phAE41; Escherichia coli bacteriophage cosmid;
KW   pYUB328; mycobacteria; mutation; vaccine; treatment;
KW   mycobacterial infection; ss.
OS   Chimeric - Bacteriophage L5.
OS   Chimeric - Bacteriophage cosmid pYUB328.
PN   US5750384-A.
PD   12-MAY-1998.
PF   23-MAY-1994; 247901.
PR   23-MAY-1994; US-247901.
PR   07-FEB-1992; US-833431.
PR   29-APR-1993; US-057531.
PA   (UYPI-) UNIV PITTSBURGH.
PA   (YESH ) UNIV YESHIVA EINSTEIN COLLEGE.
PI   Bardarov S, Hatfull GF, Jacobs WR, McAdam R;
DR   WPI; 98-296771/26.
PT   Mycobacterial shuttle plasmids comprising mycobacteriophage L5 with
PT   lambda cosmid insert - useful for delivery of foreign DNA into
PT   mycobacteria and for generating mutations
PS   Claim 7; Columns 5-28; 42pp; English.
CC   The present sequence represents a L5 shuttle phasmid designated phAE41.
CC   It was prepared by inserting an Escherichia coli bacteriophage cosmid,
CC   pYUB328, into a non-essential region of an L5 mycobacteriophage genome.
CC   Foreign DNA such as reporter genes (e.g. luciferase), transposons
CC   (e.g. IS1096) and mycobacterial inhibiter genes can also be inserted
CC   into the phasmid. The phasmids can be used to introduce foreign DNA
CC   into mycobacteria. The L5 shuttle phasmid can be used to introduce
CC   mutations that can be used to study the mechanisms of mycobacteria and
CC   to develop vaccines and drugs for treating mycobacterial infections.
SQ   Sequence   50341 BP;    9641 A;    15673 C;    15276 G;    9750 T;
```

alignment_scores:

```
            Quality:  108.50            Length:    251
              Ratio:  0.719               Gaps:     12
 Percent Similarity:  60.159   Percent Identity: 22.709
```

alignment_block:

US-09-068-804-5 x V22674 ..

Align seg 1/1 to: V22674 from: 1 to: 50341

```
     13 IleGlyLysValLeuGlyAlaLeuLeuThrIleValSerValValAlaAl 29
  16998 ATCGGCAACGTCCTCGGGACGCTGGGCACACAGCTCGCTCCCATCGTCAC 17047

     29 aValPheThrGlyGlyAlaSerLeuAlaLeuAlaAlaValGlyLeuAlaV 46
  17048 GGCGCTCACG......CCGGCCTTCCAGACGCTGGCGAGCACGCTCGGCA 17091

     46 alMetValAlaAspGluIleValLysAlaAlaThrGlyValSerPheIle 62
  17092 CGATGCTCACCGGAGCCCTC..... 17111

     63 GlnGlnAlaLeuAsnProIleMetGluHisValLeuLysProLeuMetGl 79
  17112 ...CAAGCTCTGGGTCCGATCCTGACTCAGGTC.....GCTAC 17146

     79 uLeuIleGlyLysAlaIleThrLysAlaLeuGluGlyLeu..... 92
  17147 GTTGATCGGCACGACGCTGAACACGGCGCTGCAGGCTCTCCAGCCGATGC 17196

     93 .....GlyValAspArgLysArgGlnArgTrpProAlaAlaLeuLeuVal 107
  17197 TGCCGTCGCTCATGCAGAGCTTCCAGCAGATCTCCGACGTACTGGTGACC 17246

    108 ArgLeuSerProLeuCysHisGlyAspAlaValIleValValValAlaVa 124
  17247 AGTCTGGCCCCG......CACATCCCGGCGCTGGCGACGGCCCTCGGCCA 17290

    124 lValGlyLysGlyAlaAlaAlaLysLeuGlyAsnAlaLeuSerLysMetM 141
  17291 GGTCGCA...GGCGCGGTGCTGCAGCTCGCTCCGACGATCATCTCGACGT 17337

    141 etGlyGluThrIleLysLysLeuValProAsnVal..... 152
  17338 TGGTTCCGGCGTTCGTTCAGTTGGTCCCAAAGGTCGCTGAGCTAGTTCCG 17387

    153 ...LeuLysGlnLeuAlaGlnAsnGlySerLysLeu..... 163
  17388 ACCATCGTCAACCTGGTCCAGTCGTTCGCCAACCTGATGCCGGTGGTTCT 17437

    164 ....PheThrGlnGlyMetGlnArgIleThrSerGlyLeuGlyAsnValG 179
  17438 GCCCCTGGCGCAGGCTCTGGTCAGCGTTGCTGGCGCGGTGATTCAGGTGG 17487

    179 lySerLysMet...GlyLeuGlnThrAsnAlaLeuSerLys.....Glu 192
  17488 GTGTCTCCATCGGCGGCGCGCTCATCGGCGCGCTGGCGAACCTCACGGAG 17537

    193 LeuValGlyAsnThrLeuAsnLysVal..... 201
```

```
          ::::::::::|||::::::::::||||||
   17538 ATCATCTCCAACGTCATCAAGAAGGTGTCCGAGTGGGTCAGCAGCTTCTC 17587

     202 .AlaLeuGlyMetGluValThrAsnThrAlaAlaGlnSerAlaGlyGlyV 218
          :::   :::   ::::::::::::::::||||||:::   :::|||   :
   17588 CAGCGGAGCCCAGCAGATCGCTGCGAAGGCAGCGGAACTGCCGGGGATGA 17637

     218 alAlaGluGlyValPheIleLysAsnAlaSerGluAlaLeuAlaAspPhe 234
          ::                                |||||||||:::::
   17638 TC.........................CAGTCGGCTCTCGCCAACCTG 17660

     235 Met 235
          |||
   17661 ATG 17663

seq_name: N_Geneseq_36:T51411

seq_documentation_block:
ID   T51411 standard; DNA; 52297 BP.
AC   T51411;
DT   23-APR-1997  (first entry)
DE   Mycobacteriophage L5 genome sequence.
KW   Reporter mycobacteriophage; Mycobacterium tuberculosis; infection;
KW   drug susceptibility; diagnosis; ss.
OS   Mycobacteriophage L5.
PN   WO9425572-A1.
PD   10-NOV-1994.
PF   29-APR-1994; U04788.
PR   29-APR-1993; US-057531.
PA   (UYPI-) UNIV PITTSBURGH.
PA   (YESH ) UNIV YESHIVA EINSTEIN COLLEGE.
PI   Bloom BR,  Hatfull GF,  Jacobs WR;
DR   WPI; 95-022252/03.
PT   Reporter myco-bacteriophages - useful for rapid diagnosis of
PT   myco-bacterial infection and assessment of drug susceptibilities to
PT   mycobacterial strains.
PS   Disclosure; Page 103-137; 192pp; English.
CC   The entire genome sequence of mycobacteriophage L5 is given in
CC   T51411.  L5 is a temperate virus with broad host range among
CC   mycobacteria.  Transcriptional promoters and reporter genes (esp.
CC   luciferase or beta-galactosidase) may be introduced into the genome
CC   sequence to provide mycobacterial species-specific reporter
CC   mycobacteriophages, e.g. phGS1 (ATCC 75454) and phGS5 (ATCC 75453).
CC   These are useful in the rapid diagnosis of mycobacterial infection
CC   and for the assessment of drug susceptibilities of mycobacteria in
CC   clinical samples.  Specifically, a reporter mycobacteriophage can
CC   be used to diagnose tuberculosis and to assess the drug
CC   susceptbilities of various strains of Mycobacterium tuberculosis.
SQ   Sequence   52297 BP;    9633 A;    16532 C;    16025 G;    10107 T;


alignment_scores:
           Quality:  108.50              Length:    251
             Ratio:  0.719                Gaps:     12
 Percent Similarity:  60.159   Percent Identity: 22.709

alignment_block:
 US-09-068-804-5 x T51411   ..

 Align seg 1/1  to: T51411  from: 1  to: 52297

      13 IleGlyLysValLeuGlyAlaLeuLeuThrIleValSerValValAlaAl 29
         ||||||:::|||||||||:::|||   |||   ::::::   ::::::::
   16895 ATCGGCAACGTCCTCGGGACGCTGGGCACACAGCTCGCTCCCATCGTCAC 16944

      29 aValPheThrGlyGlyAlaSerLeuAlaLeuAlaAlaValGlyLeuAlaV 46
         ::::::::|||      :::::::::         |||      |||::::
   16945 GGCGCTCACG......CCGGCCTTCCAGACGCTGGCGAGCACGCTCGGCA 16988

      46 alMetValAlaAspGluIleValLysAlaAlaThrGlyValSerPheIle 62
         ::|||:::::::::::::::
   16989 CGATGCTCACCGGAGCCCTC.............................. 17008

      63 GlnGlnAlaLeuAsnProIleMetGluHisValLeuLysProLeuMetGl 79
            |||||||||:::||||||:::   :::|||
   17009 ...CAAGCTCTGGGTCCGATCCTGACTCAGGTC............GCTAC 17043

      79 uLeuIleGlyLysAlaIleThrLysAlaLeuGluGlyLeu..... 92
         |||||||||:::::::::::::::||||||:::::|||
   17044 GTTGATCGGCACGACGCTGAACACGGCGCTGCAGGCTCTCCAGCCGATGC 17093

      93 .....GlyValAspArgLysArgGlnArgTrpProAlaAlaLeuLeuVal 107
           ::::::   :::      |||:::   ::::::::::|||::::::
   17094 TGCCGTCGCTCATGCAGAGCTTCCAGCAGATCTCCGACGTACTGGTGACC 17143

     108 ArgLeuSerProLeuCysHisGlyAspAlaValIleValValValAlaVa 124
         |||:::|||      |||      |||:::   ::::::::::::
   17144 AGTCTGGCCCCG......CACATCCCGGCGCTGGCGACGGCCCTCGGCCA 17187

     124 lValGlyLysGlyAlaAlaAlaLysLeuGlyAsnAlaLeuSerLysMetM 141
          |||:::   ||||||:::   :::|||:::   ::::::         :
   17188 GGTCGCA...GGCGCGGTGCTGCAGCTCGCTCCGACGATCATCTCGACGT 17234

     141 etGlyGluThrIleLysLysLeuValProAsnVal..... 152
         ::      ::::::   :::||||||||||:::|||
   17235 TGGTTCCGGCGTTCGTTCAGTTGGTCCCAAAGGTCGCTGAGCTAGTTCCG 17284

     153 ...LeuLysGlnLeuAlaGlnAsnGlySerLysLeu..... 163
            :::   :::|||:::|||:::   ::::::|||
   17285 ACCATCGTCAACCTGGTCCAGTCGTTCGCCAACCTGATGCCGGTGGTTCT 17334

     164 ....PheThrGlnGlyMetGlnArgIleThrSerGlyLeuGlyAsnValG 179
             :::::::|||::::::      :::::::::::::   :::||||
   17335 GCCCCTGGCGCAGGCTCTGGTCAGCGTTGCTGGCGCGGTGATTCAGGTGG 17384

     179 lySerLysMet...GlyLeuGlnThrAsnAlaLeuSerLys......Glu 192
         ||      :::   |||      ::::::||||||::::::      |||
   17385 GTGTCTCCATCGGCGGCGCGCTCATCGGCGCGCTGGCGAACCTCACGGAG 17434

     193 LeuValGlyAsnThrLeuAsnLysVal..... 201
         :::::::::|||::::::::::|||||||
   17435 ATCATCTCCAACGTCATCAAGAAGGTGTCCGAGTGGGTCAGCAGCTTCTC 17484

     202 .AlaLeuGlyMetGluValThrAsnThrAlaAlaGlnSerAlaGlyGlyV 218
          :::   :::   ::::::::::::::::||||||:::   :::|||   :
   17485 CAGCGGAGCCCAGCAGATCGCTGCGAAGGCAGCGGAACTGCCGGGGATGA 17534

     218 alAlaGluGlyValPheIleLysAsnAlaSerGluAlaLeuAlaAspPhe 234
          ::                                |||||||||:::::
   17535 TC.........................CAGTCGGCTCTCGCCAACCTG 17557

     235 Met 235
          |||
   17558 ATG 17560

seq_name: N_Geneseq_36:Q47357

seq_documentation_block:
ID   Q47357 standard; DNA; 52298 BP.
AC   Q47357;
DT   07-JUN-1994  (first entry)
DE   L5 mycobacteriophage DNA.
KW   Genome; L5; mycobacteriophage; temperate virus; mycobacteria; phage;
KW   phage g; lysogeny; integration; chromosome; attP; attachment site;
KW   62-61-60 gene; left arm; integrase; integration-proficient vector;
KW   reporter gene; diagnosis; right arm; tuberculosis; drug resistance;
KW   mycobacterial species-specific reporter mycobacteriophage; ss.
OS   L5 mycobacteriophage.
PN   WO9316172-A.
PD   19-AUG-1993.
PF   02-FEB-1993; U00913.
PR   07-FEB-1992; US-833431.
PA   (UYPI-) UNIV PITTSBURGH.
PA   (YESH ) UNIV YESHIVA EINSTEIN COLLEGE.
PI   Bloom BR,  Hatfull GF,  Jacobs WR;
DR   WPI; 93-272877/34.
```

```
PT   Mycobacterial species-specific reporter myco-bacteriophages -
PT   used to detect infection by or drug resistance of mycobacteria
PS   Disclosure; Page 11-25; 65pp; English.
CC   This sequence represents the DNA sequence of the entire genome of
CC   the L5 mycobacteriophage.  L5 is a temperate virus with a broad host
CC   range among mycobacteria, and is morphologically similar to the
CC   family of phages that includes phage g and contain a linear double
CC   stranded genome with cohesive ends.  During the establishment of
CC   lysogeny, the L5 genome becomes integrated into the mycobacterial
CC   chromosome at the attP attachment site.  DNA analysis has indicated
CC   that the L5 genome is organised into a right and left arm with the
CC   attachment site and integrase at the center of the genome.  The L5
CC   genome may be used for the production of integration-proficient
CC   vectors for mycobacteria.  Part of the L5 genome is not essential for
CC   mycobateriophage growth.  It has been demonstrated that all or part of
CC   the gene 62-61-60 can be deleted without effecting the life cycle of
CC   the L5 phage.  This region is therefore suitable for insertion of
CC   reporter genes.  Mycobacterial species-specific reporter mycobacterio-
CC   phages produced in this way may be used to diagnose mycobacterial
CC   disease, preferably tuberculosis, by incubating with a sample and
CC   detecting the reporter gene product.  They may also be used for
CC   assessing drug resistance of a mycobacterial strain by adding the
CC   reporter vector to the strain, then adding the drug and detecting
CC   the formation of any gene product.
SQ   Sequence   52298 BP;    9631 A;    16531 C;    16022 G;    10106 T;


alignment_scores:
           Quality:  108.50              Length:    251
             Ratio:  0.719                Gaps:     12
 Percent Similarity:  60.159   Percent Identity: 22.709

alignment_block:
 US-09-068-804-5 x Q47357   ..

 Align seg 1/1  to: Q47357  from: 1  to: 52298

     13 IleGlyLysValLeuGlyAlaLeuLeuThrIleValSerValValAlaAl 29
        ||||||:::|||||||||:::|||   |||   ::::::   ::::::::
  16895 ATCGGCAACGTCCTCGGGACGCTGGGCACACAGCTCGCTCCCATCGTCAC 16944

     29 aValPheThrGlyGlyAlaSerLeuAlaLeuAlaAlaValGlyLeuAlaV 46
        :::::::|||      :::::::::         |||      |||::::
  16945 GGCGCTCACG......CCGGCCTTCCAGACGCTGGCGAGCACGCTCGGCA 16988

     46 alMetValAlaAspGluIleValLysAlaAlaThrGlyValSerPheIle 62
        ::|||:::::::::::::::
  16989 CGATGCTCACCGGAGCCCTC............................. 17008

     63 GlnGlnAlaLeuAsnProIleMetGluHisValLeuLysProLeuMetGl 79
           |||||||||:::|||||||:::   :::|||
  17009 ...CAAGCTCTGGGTCCGATCCTGACTCAGGTC............GCTAC 17043

     79 uLeuIleGlyLysAlaIleThrLysAlaLeuGluGlyLeu.......... 92
         |||||||||:::::::::::::::::|||||||:::::|||
  17044 GTTGATCGGCACGACGCTGAACACGGCGCTGCAGGCTCTCCAGCCGATGC 17093

     93 .....GlyValAspArgLysArgGlnArgTrpProAlaAlaLeuLeuVal 107
          ::::::   :::      |||:::   ::::::::::|||::::::
  17094 TGCCGTCGCTCATGCAGAGCTTCCAGCAGATCTCCGACGTACTGGTGACC 17143

    108 ArgLeuSerProLeuCysHisGlyAspAlaValIleValValValAlaVa 124
           |||:::|||      |||      |||:::   ::::::::::::
  17144 AGTCTGGCCCCG......CACATCCCGGCGCTGGCGACGGCCCTCGGCCA 17187

    124 lValGlyLysGlyAlaAlaAlaLysLeuGlyAsnAlaLeuSerLysMetM 141
         |||:::   ||||||:::   :::|||:::   ::::::         :
  17188 GGTCGCA...GGCGCGGTGCTGCAGCTCGCTCCGACGATCATCTCGACGT 17234

    141 etGlyGluThrIleLysLysLeuValProAsnVal.............. 152
        ::      ::::::   :::||||||||||:::|||
  17235 TGGTTCCGGCGTTCGTTCAGTTGGTCCCAAAGGTCGCTGAGCTAGTTCCG 17284

    153 ...LeuLysGlnLeuAlaGlnAsnGlySerLysLeu.............. 163
           :::   :::|||:::|||:::   ::::::|||
  17285 ACCATCGTCAACCTGGTCCAGTCGTTCGCCAACCTGATGCCGGTGGTTCT 17334

    164 ....PheThrGlnGlyMetGlnArgIleThrSerGlyLeuGlyAsnValG 179
            ::::::|||::::::      ::::::::::::::   :::||||
  17335 GCCCCTGGCGCAGGCTCTGGTCAGCGTTGCTGGCGCGGTGATTCAGGTGG 17384

    179 lySerLysMet...GlyLeuGlnThrAsnAlaLeuSerLys......Glu 192
        ||      :::   |||      ::::::||||||::::::      |||
  17385 GTGTCTCCATCGGCGGCGCGCTCATCGGCGCGCTGGCGAACCTCACGGAG 17434

    193 LeuValGlyAsnThrLeuAsnLysVal....................... 201
        :::::::::|||::::::::::||||||
  17435 ATCATCTCCAACGTCATCAAGAAGGTGTCCGAGTGGGTCAGCAGCTTCTC 17484

    202 .AlaLeuGlyMetGluValThrAsnThrAlaAlaGlnSerAlaGlyGlyV 218
         :::   :::   ::::::::::::::::|||||||:::   :::|||   :
  17485 CAGCGGAGCCCAGCAGATCGCTGCGAAGGCAGCGGAACTGCCGGGGATGA 17534

    218 alAlaGluGlyValPheIleLysAsnAlaSerGluAlaLeuAlaAspPhe 234
        ::                                  |||||||||:::::::
  17535 TC.........................CAGTCGGCTCTCGCCAACCTG 17557

    235 Met 235
        |||
  17558 ATG 17560

seq_name: N_Geneseq_36:T64548

seq_documentation_block:
ID   T64548 standard; cDNA; 1473 BP.
AC   T64548;
DT   10-JUL-1997  (first entry)
DE   NADP-specific glutamate dehydrogenase beta subunit cDNA.
KW   Glutamate dehydrogenase; NADP-GDH; alga; nitrogen metabolism;
KW   chloroplast; transgenic plant; ds.
OS   Chlorella sorokiniana strain UTEX 1230.
FH   Key             Location/Qualifiers
FT   cds             4. .1434
FT                   /*tag=  a
FT   mat_peptide     4. .1431
FT                   /*tag=  b
PN   WO9712983-A1.
PD   10-APR-1997.
PF   03-OCT-1996; U15921.
PR   06-OCT-1995; US-541033.
PA   (UYFL ) UNIV FLORIDA.
PI   Miller P,  Schmidt RR;
DR   WPI; 97-226226/20.
DR   P-PSDB; W15412.
PT   DNA encoding NADP-GDH alpha and beta sub-units  - isolated from
PT   Chlorella sorokiniana algal protoplast(s), useful to modulate
PT   nitrogen metabolism plant cells
PS   Claim 2; Page 48-50; 61pp; English.
CC   cDNA clones (T64547 and T64548) respectively code for the alpha
CC   (W15411) and beta mature subunits (W15412) of ammonium-inducible,
CC   chloroplast-localised NADP-specific glutamate dehydrogenase (NADP-
CC   GDH) hexameric isoenzymes of Chlorella sorokiniana.  They were
CC   obtd. by removal of the chloroplast targeting signal from full-
CC   length cDNA clones (see also T64542-43) by PCR mutagenesis.  The N
CC   metabolism of plant cells can be modulated (pref. increasing the
CC   assimilation of inorganic N into organic N) by transforming them
CC   with nucleotide sequences encoding the alpha and/or beta mature
CC   subunits or subunit precursors (see also T64529-30).  Such plants
CC   show improved properties, e.g. increased crop yield and improved
CC   stress tolerance.  Heterohexamers having alpha and beta subunits
CC   can be expressed that have higher aminating:deaminating activity
CC   ratios (i.e. higher capacity for glutamate synthesis) than
CC   homohexamers.
SQ   Sequence   1473 BP;    297 A;    486 C;    470 G;    220 T;
```

```
alignment_scores:
            Quality:   91.00              Length:    267
              Ratio:  0.615                Gaps:     15
 Percent Similarity:  55.431   Percent Identity: 21.348

alignment_block:
 US-09-068-804-5 x T64548/rev   ..

 Align seg 1/1  to reverse of: T64548  from: 1  to: 1473

      17 LeuGlyAlaLeuLeuThrIleValSerValValAlaAlaValPheThrGl 33
         |||::::::   |||:::   :::::::::||||||   :::      ||
    1035 CTGGCAGTCCAGCTCCCAAGGCTTGCGGCGGTCGCCCACATA...CACGG 989

      33 yGlyAlaSerLeuAlaLeuAlaAlaValGlyLeuAlaValMetValAlaA 50
         ||||::::::|||:::|||:::::::::|||   :::|||::::::   :
     988 CGGTGTCGCTCTTGTACTCGGAGATGCGGGCGCTGTTGTTCTTCTTCTTC 939

      50 spGluIleValLys............AlaAlaThrGlyValSerPheIle 62
         ::   :::   :::            |||      :::|||:::::::::
     938 ATGTCCTGCACCGCCTGCAGCTGCTCGCGCGTGAAGCCGTTGGGCTCGTA 889

      63 GlnGlnAlaLeu.......................AsnProIleMetGl 71
         :::   ||||||                       :::::::::::::
     888 CACGTAGCCCTGGGAGTCGGACAGCGACAGCACGATGGCGCCCTTCTCCA 839

      71 uHisValLeuLysProLeuMetGluLeuIleGlyLysAla.........I 85
         ::::   |||:::::::::::      :::::::::|||
     838 GCAGCAGCTCCGCGCAGTACTGGGCCACGTTGCCCGCGCCAGACACCAGG 789

      85 leThrLysAlaLeuGluGlyLeuGlyVal..................... 94
         :::   |||||||||:::|||:::::
     788 CAGCGCTTGCCCTTGAGGCTCTCGCCCTTGTCCTTCAGCACGTTCTCCAC 739

      95 ...Asp..ArgLysArgGlnArgTrpPro.AlaAlaLeuLeuValArgLe 109
            :::  |||:::|||   |||||||||  :::|||   |||
     738 AAACAGCACGGCGCCGTAGCCGGTGGCCTCGGGGCGGATCTC........ 697

     109 uSerProLeuCysHisGlyAspAlaValIleValValValAlaValValG 126
                        ::::::   ::::::::::|||:::   |
     696 .....................GGAGCCGCCATACTCCTGGCCCTTCGGG 669

     126 lyLysGlyAlaAlaAlaLysLeuGlyAsnAlaLeuSerLysMetMetGly 142
         ||:::   |||::::::   ||||||:::||||||      :::::::::
     668 GTCAGCACGCCGGTGTAGTTCTTGGTGATGCGCTT......GTACTGGCC 625

     143 GluThrIleLysLysLeuValProAsnValLeuLysGlnLeuAlaGlnAs 159
         |||::::::   :::|||
     624 GAAAAGGTAGCCAATCTC............................... 607

     159 nGlySerLysLeuPheThrGlnGlyMetGlnArgIleThrSerGlyLeuG 176
                                              :::  :::   :
     606 ....................................GCGCGCGCCCACG 594

     176 lyAsnValGlySerLysMetGlyLeuGlnThrAsnAlaLeuSerLysGlu 192
         :::::|||::::::      |||:::::::::::::::::   :::
     593 CCGATGTCGCCGGCGGGCACGTCCTGCACGTAGCTGATGTGGCGCTGCAG 544

     193 LeuValGlyAsnThrLeuAsn..............LysValAlaLeuGl 204
         |||      :::   |||:::               :::||||||||::
     543 CTCGGTCATGAAGGACTGGCAGAAGCGCATCACCTCCGCGTCGCTCTTGC 494

     204 yMetGluValThrAsnThrAlaAlaGlnSerAlaGly...GlyValAlaG 220
         ::::   |||         ||||||   :::|||:::   |||   :::
     493 CCTTGGGGTCGAAGTCGGAGCCGCCCTTGCCGCCGCCCATGGGCAGGGTG 444

     220 luGly.ValPheIleLysAsnAlaSerGluAlaLeuAlaAspPheMetLe 236
               :::|||:::|||      |||  |||      :::|||||::
     443 GTCAGGCTGTTCTTGAAGATCTGCTCAAAGGCA...AGGAACTTCATGAT 397

     236 uAlaArgPheAlaMetAspGlnIleGlnGlnTrpLeuLysGlnSer 251
         :::::||||||:::            :::   |||   :::::::::
     396 GGACAGGTTCACG............GAGGGGTGGAAGCGCAGGCCG 363
```

```
seq_name: N_Geneseq_36:T64547

seq_documentation_block:
ID   T64547 standard; cDNA; 1506 BP.
AC   T64547;
DT   10-JUL-1997  (first entry)
DE   NADP-specific glutamate dehydrogenase alpha subunit cDNA.
KW   Glutamate dehydrogenase; NADP-GDH; alga; nitrogen metabolism;
KW   chloroplast; transgenic plant; ds.
OS   Chlorella sorokiniana strain UTEX 1230.
FH   Key             Location/Qualifiers
FT   cds             4. .1467
FT                   /*tag=  a
FT   mat_peptide     4. .1464
FT                   /*tag=  b
PN   WO9712983-A1.
PD   10-APR-1997.
PF   03-OCT-1996; U15921.
PR   06-OCT-1995; US-541033.
PA   (UYFL ) UNIV FLORIDA.
PI   Miller P,  Schmidt RR;
DR   WPI; 97-226226/20.
DR   P-PSDB; W15411.
PT   DNA encoding NADP-GDH alpha and beta sub-units  - isolated from
PT   Chlorella sorokiniana algal protoplast(s), useful to modulate
PT   nitrogen metabolism plant cells
PS   Claim 2; Page 44-46; 61pp; English.
CC   cDNA clones (T64547 and T64548) respectively code for the alpha
CC   (W15411) and beta mature subunits (W15412) of ammonium-inducible,
CC   chloroplast-localised NADP-specific glutamate dehydrogenase (NADP-
CC   GDH) hexameric isoenzymes of Chlorella sorokiniana.  They were
CC   obtd. by removal of the chloroplast targeting signal from full-
CC   length cDNA clones (see also T64542-43) by PCR mutagenesis.  The N
CC   metabolism of plant cells can be modulated (pref. increasing the
CC   assimilation of inorganic N into organic N) by transforming them
CC   with nucleotide sequences encoding the alpha and/or beta mature
CC   subunits or subunit precursors (see also T64529-30).  Such plants
CC   show improved properties, e.g. increased crop yield and improved
CC   stress tolerance.  Heterohexamers having alpha and beta subunits
CC   can be expressed that have higher aminating:deaminating activity
CC   ratios (i.e. higher capacity for glutamate synthesis) than
CC   homohexamers.
SQ   Sequence   1506 BP;    302 A;    495 C;    483 G;    226 T;
```

```
alignment_scores:
            Quality:   91.00              Length:    267
              Ratio:  0.615                Gaps:     15
 Percent Similarity:  55.431   Percent Identity: 21.348

alignment_block:
 US-09-068-804-5 x T64547/rev   ..

 Align seg 1/1  to reverse of: T64547  from: 1  to: 1506

      17 LeuGlyAlaLeuLeuThrIleValSerValValAlaAlaValPheThrGl 33
         |||::::::   |||:::   :::::::::||||||   :::      ||
    1068 CTGGCAGTCCAGCTCCCAAGGCTTGCGGCGGTCGCCCACATA...CACGG 1022

      33 yGlyAlaSerLeuAlaLeuAlaAlaValGlyLeuAlaValMetValAlaA 50
         ||||::::::|||:::|||:::::::::|||   :::|||::::::   :
    1021 CGGTGTCGCTCTTGTACTCGGAGATGCGGGCGCTGTTGTTCTTCTTCTTC 972

      50 spGluIleValLys............AlaAlaThrGlyValSerPheIle 62
         ::   :::   :::            |||      :::|||:::::::::
     971 ATGTCCTGCACCGCCTGCAGCTGCTCGCGCGTGAAGCCGTTGGGCTCGTA 922

      63 GlnGlnAlaLeu.......................AsnProIleMetGl 71
         :::   ||||||                       :::::::::::::
     921 CACGTAGCCCTGGGAGTCGGACAGCGACAGCACGATGGCGCCCTTCTCCA 872

      71 uHisValLeuLysProLeuMetGluLeuIleGlyLysAla.........I 85
```

```
     :::   |||:::::::::::::      ::::::::::|||
871 GCAGCAGCTCCGCGCAGTACTGGGCCACGTTGCCCGCGCCAGACACCAGG 822

 85 leThrLysAlaLeuGluGlyLeuGlyVal..................... 94
      :::   |||||||||:::|||:::::::
821 CAGCGCTTGCCCTTGAGGCTCTCGCCCTTGTCCTTCAGCACGTTCTCCAC 772

 95 ...Asp..ArgLysArgGlnArgTrpPro.AlaAlaLeuLeuValArgLe 109
       :::  |||:::|||   |||||||||| :::|||   |||
771 AAACAGCACGGCGCCGTAGCCGGTGGCCTCGGGGCGGATCTC........ 730

109 uSerProLeuCysHisGlyAspAlaValIleValValValAlaValValG 126
                      ::::::   ::::::::::|||:::   |
729 .....................GGAGCCGCCATACTCCTGGCCCTTCGGG 702

126 lyLysGlyAlaAlaAlaLysLeuGlyAsnAlaLeuSerLysMetMetGly 142
    ||:::   |||::::::   ||||||:::|||||||       :::::::::
701 GTCAGCACGCCGGTGTAGTTCTTGGTGATGCGCTT......GTACTGGCC 658

143 GluThrIleLysLysLeuValProAsnValLeuLysGlnLeuAlaGlnAs 159
    |||::::::   :::|||
657 GAAAAGGTAGCCAATCTC................................ 640

159 nGlySerLysLeuPheThrGlnGlyMetGlnArgIleThrSerGlyLeuG 176
                                        :::   :::   :
639 ....................................GCGCGCGCCCACG 627

176 lyAsnValGlySerLysMetGlyLeuGlnThrAsnAlaLeuSerLysGlu 192
    :::::|||:::::::         |||:::::::::::::::::::   :::
626 CCGATGTCGCCGGCGGGCACGTCCTGCACGTAGCTGATGTGGCGCTGCAG 577

193 LeuValGlyAsnThrLeuAsn..............LysValAlaLeuGl 204
    |||      :::   |||:::              :::||||||||||::
576 CTCGGTCATGAAGGACTGGCAGAAGCGCATCACCTCCGCGTCGCTCTTGC 527

204 yMetGluValThrAsnThrAlaAlaGlnSerAlaGly...GlyValAlaG 220
    ::::   |||         ||||||   :::|||:::   |||   :::
526 CCTTGGGGTCGAAGTCGGAGCCGCCCTTGCCGCCGCCCATGGGCAGGGTG 477

220 luGly.ValPheIleLysAsnAlaSerGluAlaLeuAlaAspPheMetLe 236
          :::|||:::|||      |||   |||      :::|||||::
476 GTCAGGCTGTTCTTGAAGATCTGCTCAAAGGCA...AGGAACTTCATGAT 430

236 uAlaArgPheAlaMetAspGlnIleGlnGlnTrpLeuLysGlnSer 251
    ::::||||||:::           :::   |||   :::::::::
429 GGACAGGTTCACG............GAGGGGTGGAAGCGCAGGCCG 396
```

seq_name: N_Geneseq_36:T64531

seq_documentation_block:

ID T64531 standard; cDNA; 1969 BP.
AC T64531;
DT 10-JUL-1997 (first entry)
DE Glutamate dehydrogenase cDNA clone pBGDc53.
KW Glutamate dehydrogenase; NADP-GDH; alga; nitrogen metabolism;
KW chloroplast; transgenic plant; ds.
OS Chlorella sorokiniana strain UTEX 1230.
PN WO9712983-A1.
PD 10-APR-1997.
PF 03-OCT-1996; U15921.
PR 06-OCT-1995; US-541033.
PA (UYFL) UNIV FLORIDA.
PI Miller P, Schmidt RR;
DR WPI; 97-226226/20.
PT DNA encoding NADP-GDH alpha and beta sub-units - isolated from
PT Chlorella sorokiniana algal protoplast(s), useful to modulate
PT nitrogen metabolism plant cells
PS Claim 2; Page 35-36; 61pp; English.
CC A cDNA clone (T64531), designated pBGDc53, was isolated from a
CC Chlorella sorokiniana cDNA library in lambda ZAP by screening
CC with an NADP-specific glutamate dehydrogenase (NADP-GDH) cDNA
CC probe. It comprises the complete 3'-terminal region of the
CC C. sorokiniana NADP-GDH sequence. 5' terminal regions were
CC obtd. by PCR, yielding full-length clones (T64542-43) for
CC the alpha and beta subunits of C. sorokiniana NADP-GDH.
SQ Sequence 1969 BP; 401 A; 659 C; 580 G; 329 T;

alignment_scores:
Quality: 91.00 Length: 267
Ratio: 0.615 Gaps: 15
Percent Similarity: 55.431 Percent Identity: 21.348

alignment_block:
US-09-068-804-5 x T64531/rev ..

Align seg 1/1 to reverse of: T64531 from: 1 to: 1969

```
 17 LeuGlyAlaLeuLeuThrIleValSerValValAlaAlaValPheThrGl 33
    |||:::::::   |||:::   ::::::::::||||||   :::      ||
1044 CTGGCAGTCCAGCTCCCAAGGCTTGCGGCGGTCGCCCACATA...CACGG 998

 33 yGlyAlaSerLeuAlaLeuAlaAlaValGlyLeuAlaValMetValAlaA 50
    ||||::::::|||:::|||::::::::::|||   :::|||::::::   :
997 CGGTGTCGCTCTTGTACTCGGAGATGCGGGCGCTGTTGTTCTTCTTCTTC 948

 50 spGluIleValLys.............AlaAlaThrGlyValSerPheIle 62
    ::   :::   :::              |||      :::|||::::::::::
947 ATGTCCTGCACCGCCTGCAGCTGCTCGCGCGTGAAGCCGTTGGGCTCGTA 898

 63 GlnGlnAlaLeu.........................AsnProIleMetGl 71
    :::   ||||||                        ::::::::::::::
897 CACGTAGCCCTGGGAGTCGGACAGCGACAGCACGATGGCGCCCTTCTCCA 848

 71 uHisValLeuLysProLeuMetGluLeuIleGlyLysAla.........I 85
    ::::   |||:::::::::::      ::::::::::|||
847 GCAGCAGCTCCGCGCAGTACTGGGCCACGTTGCCCGCGCCAGACACCAGG 798

 85 leThrLysAlaLeuGluGlyLeuGlyVal..................... 94
      :::   |||||||||:::|||:::::::
797 CAGCGCTTGCCCTTGAGGCTCTCGCCCTTGTCCTTCAGCACGTTCTCCAC 748

 95 ...Asp..ArgLysArgGlnArgTrpPro.AlaAlaLeuLeuValArgLe 109
       :::  |||:::|||   |||||||||| :::|||   |||
747 AAACAGCACGGCGCCGTAGCCGGTGGCCTCGGGGCGGATCTC........ 706

109 uSerProLeuCysHisGlyAspAlaValIleValValValAlaValValG 126
                      ::::::   ::::::::::|||:::   |
705 .....................GGAGCCGCCATACTCCTGGCCCTTCGGG 678

126 lyLysGlyAlaAlaAlaLysLeuGlyAsnAlaLeuSerLysMetMetGly 142
    ||:::   |||::::::   ||||||:::|||||||       :::::::::
677 GTCAGCACGCCGGTGTAGTTCTTGGTGATGCGCTT......GTACTGGCC 634

143 GluThrIleLysLysLeuValProAsnValLeuLysGlnLeuAlaGlnAs 159
    |||::::::   :::|||
633 GAAAAGGTAGCCAATCTC................................ 616

159 nGlySerLysLeuPheThrGlnGlyMetGlnArgIleThrSerGlyLeuG 176
                                        :::   :::   :
615 ....................................GCGCGCGCCCACG 603

176 lyAsnValGlySerLysMetGlyLeuGlnThrAsnAlaLeuSerLysGlu 192
    :::::|||:::::::         |||:::::::::::::::::::   :::
602 CCGATGTCGCCGGCGGGCACGTCCTGCACGTAGCTGATGTGGCGCTGCAG 553

193 LeuValGlyAsnThrLeuAsn..............LysValAlaLeuGl 204
    |||      :::   |||:::              :::||||||||||::
552 CTCGGTCATGAAGGACTGGCAGAAGCGCATCACCTCCGCGTCGCTCTTGC 503

204 yMetGluValThrAsnThrAlaAlaGlnSerAlaGly...GlyValAlaG 220
    ::::   |||         ||||||   :::|||:::   |||   :::
502 CCTTGGGGTCGAAGTCGGAGCCGCCCTTGCCGCCGCCCATGGGCAGGGTG 453

220 luGly.ValPheIleLysAsnAlaSerGluAlaLeuAlaAspPheMetLe 236
          :::|||:::|||      |||   |||      :::|||||::
```

```
     452 GTCAGGCTGTTCTTGAAGATCTGCTCAAAGGCA...AGGAACTTCATGAT 406

     236 uAlaArgPheAlaMetAspGlnIleGlnGlnTrpLeuLysGlnSer 251
         ::::||||||:::            :::   |||   :::::::::
     405 GGACAGGTTCACG............GAGGGGTGGAAGCGCAGGCCG 372

seq_name: N_Geneseq_36:Q71390

seq_documentation_block:
ID   Q71390 standard; DNA; 4614 BP.
AC   Q71390;
DT   13-APR-1995  (first entry)
DE   Yeast 4.7 kB agglutination gene FLO1L.
KW   Yeast; agglutination; FLO1L; ss.
OS   Saccharomyces cervisiae ABXL-1D.
FH   Key             Location/Qualifiers
FT   cds             1. .4614
FT                   /*tag=  a
PN   WO9419475-A.
PD   01-SEP-1994.
PF   24-FEB-1994; J00290.
PR   26-FEB-1993; JP-038871.
PA   (PANI-) PANIMOLABORATORIO BRYGGERILABORATORIUM.
PA   (SAPB ) SAPPORO BREWERIES.
PI   Keraenen S,  Ogawa M,  Onnela M,  Penttila M,  Takata Y;
PI   Watari J;
DR   WPI; 94-294338/36.
DR   P-PSDB; R60562.
PT   New yeast agglutination genes and yeast contg. them - impart
PT   agglutination properties to facilitate removal from fermentation
PT   media
PS   Claim 4; Page 43-48; 75pp; English.
CC   The agglutination gene is called FLO1. Saccharomyces cerevisiae
CC   includes an agglutination gene of 4.7 kb (FLO1L) and an
CC   agglutination gene of 2.6 kb (FLO1S). FLO1L is the intact FLO1 gene
CC   on chromosome I, and FLO1S is the FLO1L gene with a portion of the
CC   ORF deleted in frame. FLO1L imparts a relatively strong
CC   agglutinative property to the host yeast into which it is introduced,
CC   while FLO1S imparts a weaker agglutinative property.
SQ   Sequence   4614 BP;    1357 A;    1303 C;    789 G;    1165 T;


alignment_scores:
           Quality:   95.00            Length:    194
             Ratio:  0.841              Gaps:      6
 Percent Similarity:  58.247   Percent Identity: 23.196

alignment_block:
 US-09-068-804-5 x Q71390/rev   ..

 Align seg 1/1  to reverse of: Q71390  from: 1  to: 4614

      16 ValLeuGlyAlaLeuLeuThrIleValSerValValAlaAlaValPheTh 32
         |||:::   |||:::::::::::::|||:::|||:::   ::::::   ::
    1376 GTAGTCATAGCAGTAGTGGCTGTTGTTGGTGTTCTGATGACAATGATGGT 1327

      32 rGlyGlyAlaSerLeuAlaLeuAlaAlaValGlyLeuAlaValMetValA 49
         ::::::::::::   :::|||::::::|||   :::::::::   |||:
    1326 TTCATCAGTTGGCAAACCGTTGGTACCGGTGACGGTGGTCATTTCAGTGG 1277

      49 laAspGluIleValLysAlaAlaThrGlyValSerPheIleGlnGlnAla 65
         ::   |||:::                |||||||||        :::
    1276 ATGTAGAGGTA.............AAAGTGTCGTTCCATGGCTGAGTT 1242

      66 LeuAsnProIleMetGluHisValLeuLysProLeuMetGluLeuIleGl 82
         :::        :::      |||:::     ||||||   ::::::
    1241 GTAGTCATGGCAGTAGTGGCTGTTGTTGGTGTTCTGATGACAATGATGGT 1192

      82 yLysAlaIleThrLysAlaLeuGluGlyLeuGlyValAspArgLysArgG 99
         ::::::   |||:::|||      :::   |||
    1191 CTCATCAGTTGGCAAACCATTGGTACCGGTGACTGTG............. 1155

      99 lnArgTrpProAlaAlaLeuLeuValArgLeuSerProLeuCysHisGly 115
                 :::::::::   |||   :::      |||   ||||||
    1154 ........GTCAATTCGGTAGAAGTAGAGGTAAAAGTGCTGTTCCATGGC 1113

     116 AspAlaValIleVal...ValValAlaValValGlyLysGlyAlaAlaAl 131
         ::::::|||::::::   ||||||||||||||||||
    1112 TCAGTTGTAGTTATGGCAGTAGTGGCTGTTGTTGGT.............. 1077

     131 aLysLeuGlyAsnAlaLeuSerLysMetMetGlyGluThrIleLysLys. 147
                    :::|||   :::||||||      ::::::::::::
    1076 .............GTTCTGATTACAATGATGGTTTCGTCAGTTCGTACAC 1040

     148 ..LeuValProAsnValLeuLysGlnLeuAlaGlnAsnGlySerLysLeu 163
           |||||||||   ::::::   :::   ::::::          |||:::
    1039 CATTGGTGCCAGTGACTGTGGTCAATTCGGTAGAAGTAGAGGTAAAAGTG 990

     164 PheThrGlnGlyMetGlnArgIleThrSerGlyLeuGlyAsnValGlySe 180
         :::   :::|||          :::          |||:::   ||||||
     989 CTGTTCCATGGCTCAGTTGTAGTTATGATGGTGCTAGCAGTTGTTGGAGT 940

     180 rLysMetGlyLeuGlnThrAsnAlaLeuSerLysGluLeuValGlyAsnT 197
           |||   :::   ::::::::::::::::::   ||||||      |
     939 TCTGATGACAATGACGGTTTCGTCAGTTGGAACGCCGTTGGTACCGGTGA 890

     197 hrLeuAsnLysValAlaLeuGlyMetGluVal 207
         ||         |||:::::::::::::::|||
     889 CG.........GTGGTCATTTCAGTAGATGTA 867

seq_name: N_Geneseq_36:T64543

seq_documentation_block:
ID   T64543 standard; cDNA; 2096 BP.
AC   T64543;
DT   10-JUL-1997  (first entry)
DE   NADP-specific glutamate dehydrogenase beta-subunit precursor cDNA.
KW   Glutamate dehydrogenase; NADP-GDH; alga; nitrogen metabolism;
KW   chloroplast; transgenic plant; ds.
OS   Chlorella sorokiniana strain UTEX 1230.
FH   Key             Location/Qualifiers
FT   cds             30. .1568
FT                   /*tag=  a
FT   transit_peptide 30. .146
FT                   /*tag=  b
FT   mat_peptide     147. .1565
FT                   /*tag=  c
PN   WO9712983-A1.
PD   10-APR-1997.
PF   03-OCT-1996; U15921.
PR   06-OCT-1995; US-541033.
PA   (UYFL ) UNIV FLORIDA.
PI   Miller P,  Schmidt RR;
DR   WPI; 97-226226/20.
DR   P-PSDB; W15408.
PT   DNA encoding NADP-GDH alpha and beta sub-units  - isolated from
PT   Chlorella sorokiniana algal protoplast(s), useful to modulate
PT   nitrogen metabolism plant cells
PS   Claim 2; Page 41-43; 61pp; English.
CC   cDNA clones (T64542 and T64543) respectively code for the alpha
CC   subunit (W15407) and beta subunit (W15408) precursor proteins of an
CC   ammonium-inducible, chloroplast-localised NADP-specific glutamate
CC   dehydrogenase (NADP-GDH) of Chlorella sorokiniana.  They were obtd.
CC   by ligation of 3' clone pBGDc53 (T64531) and 5' clones pGDc63
CC   (T64540) and pGDc64 (T64541).  The proteins are processed to yield
CC   mature alpha and beta subunits (see also T64547-48) that comprise
CC   the active NADP-GDH hexameric isoenzymes.  The nitrogen metabolism
CC   of plants can be modulated (pref. by increasing assimilation of
CC   inorganic N into organic N) by transforming them with nucleotide
CC   sequences encoding the alpha and/or beta subunits.  Such plants show
CC   improved properties, e.g. increased crop yield and improved stress
CC   tolerance.  Heterohexamers having alpha and beta subunits can be
CC   expressed that have higher aminating:deaminating activity ratios
CC   (i.e. higher capacity for glutamate synthesis) than homohexamers.
SQ   Sequence   2096 BP;    416 A;    711 C;    616 G;    353 T;
```

```
alignment_scores:
            Quality:   91.00              Length:    267
              Ratio:  0.615                Gaps:     15
 Percent Similarity:  55.431   Percent Identity: 21.348

alignment_block:
 US-09-068-804-5 x T64543/rev   ..

 Align seg 1/1  to reverse of: T64543  from: 1  to: 2096

      17 LeuGlyAlaLeuLeuThrIleValSerValValAlaAlaValPheThrGl 33
    1169 CTGGCAGTCCAGCTCCCAAGGCTTGCGGCGGTCGCCCACATA...CACGG 1123

      33 yGlyAlaSerLeuAlaLeuAlaAlaValGlyLeuAlaValMetValAlaA 50
    1122 CGGTGTCGCTCTTGTACTCGGAGATGCGGGCGCTGTTGTTCTTCTTCTTC 1073

      50 spGluIleValLys.....AlaAlaThrGlyValSerPheIle 62
    1072 ATGTCCTGCACCGCCTGCAGCTGCTCGCGCGTGAAGCCGTTGGGCTCGTA 1023

      63 GlnGlnAlaLeu.....AsnProIleMetGl 71
    1022 CACGTAGCCCTGGGAGTCGGACAGCGACAGCACGATGGCGCCCTTCTCCA 973

      71 uHisValLeuLysProLeuMetGluLeuIleGlyLysAla.....I 85
     972 GCAGCAGCTCCGCGCAGTACTGGGCCACGTTGCCCGCGCCAGACACCAGG 923

      85 leThrLysAlaLeuGluGlyLeuGlyVal..... 94
     922 CAGCGCTTGCCCTTGAGGCTCTCGCCCTTGTCCTTCAGCACGTTCTCCAC 873

      95 ...Asp..ArgLysArgGlnArgTrpPro.AlaAlaLeuLeuValArgLe 109
     872 AAACAGCACGGCGCCGTAGCCGGTGGCCTCGGGGCGGATCTC..... 831

     109 uSerProLeuCysHisGlyAspAlaValIleValValValAlaValValG 126
     830 .....GGAGCCGCCATACTCCTGGCCCTTCGGG 803

     126 lyLysGlyAlaAlaAlaLysLeuGlyAsnAlaLeuSerLysMetMetGly 142
     802 GTCAGCACGCCGGTGTAGTTCTTGGTGATGCGCTT......GTACTGGCC 759

     143 GluThrIleLysLysLeuValProAsnValLeuLysGlnLeuAlaGlnAs 159
     758 GAAAAGGTAGCCAATCTC..... 741

     159 nGlySerLysLeuPheThrGlnGlyMetGlnArgIleThrSerGlyLeuG 176
     740 .....GCGCGCGCCCACG 728

     176 lyAsnValGlySerLysMetGlyLeuGlnThrAsnAlaLeuSerLysGlu 192
     727 CCGATGTCGCCGGCGGGCACGTCCTGCACGTAGCTGATGTGGCGCTGCAG 678

     193 LeuValGlyAsnThrLeuAsn.....LysValAlaLeuGl 204
     677 CTCGGTCATGAAGGACTGGCAGAAGCGCATCACCTCCGCGTCGCTCTTGC 628

     204 yMetGluValThrAsnThrAlaAlaGlnSerAlaGly...GlyValAlaG 220
     627 CCTTGGGGTCGAAGTCGGAGCCGCCCTTGCCGCCGCCCATGGGCAGGGTG 578

     220 luGly.ValPheIleLysAsnAlaSerGluAlaLeuAlaAspPheMetLe 236
     577 GTCAGGCTGTTCTTGAAGATCTGCTCAAAGGCA...AGGAACTTCATGAT 531

     236 uAlaArgPheAlaMetAspGlnIleGlnGlnTrpLeuLysGlnSer 251
```

```
     530 GGACAGGTTCACG.....GAGGGGTGGAAGCGCAGGCCG 497
```

```
OM of: US-09-068-804-5 to: GenEmbl:*   out_format : pfs

Date: Oct 5, 1999  7:58 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn2_1/USPTO_spool/US09068804/runat_05101999_145134_10315/app_query.fasta.1
 -DB=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=pam150 -TRANS=human40.cdi
 -LIST=45 -DOCALIGN=200 -THR_SCORE=escore -ALIGN=15 -MODE=LOCAL
 -OUTFMT=pfs -NORM=stat -USER=US09068804 -NCPU=6 -ICPU=3 -WAIT
 -THREADS=1

Search information block:
Query: US-09-068-804-5
Query length: 290
Database: GenEmbl:*
Database sequences: 679419
Database length: 1590154680
Search time (sec): 1424.920000


score_list:
Sequence              Strd Orig    ZScore    EScore Len  ! Documentation  ..
gb_ba1:SDU66877         + 1202.50 1938.56  4.9e-100  2995  ! U66877 Salmonella dublin SicA
gb_ba2:STU30491         + 1202.50 1937.31  5.7e-100  3441  ! U30491 Salmonella typhimurium
gb_ba1:STU25631         + 1202.50 1936.85  6.1e-100  3622  ! U25631 Salmonella typhimurium
gb_ba1:STSIPGNS         + 1196.50 1923.52  3.4e-99   5393  ! X82670 S.typhi spaT, sipB, sip
gb_ba1:W10PPIBCA        +  557.00  885.02  2.4e-41   4761  ! M34849 Plasmid pWR100 (from S.
gb_ba1:SHFIPABCD        +  557.00  884.13  2.6e-41   5255  ! J04117 S.flexneri ipaBCD regic
gb_ba1:SFPVR2           +  557.00  880.74  4.1e-41   7663  ! X15319 Shigella flexineri plas
gb_ba1:SDIPABCD         +  554.00  876.84  6.7e-41   6869  ! X60777 Shigella dysenteriae ip
gb_pr3:AF061832         +  108.00  161.04  0.4975    2453  ! AF061832 Homo sapiens M4 prote
gb_pr1:HUMM4PRO         +  108.00  161.03  0.4984    2457  ! L03532 Human M4 protein mRNA,
gb_pat:A27284           +  105.00  160.55  0.5299    1506  ! A27284 H.sapiens TGR-CLH. 10/1
gb_pr2:HSTGR1           +  106.00  159.15  0.6342    2109  ! X72018 H.sapiens hTGR 1 mRNA.
gb_pat:E12520           +  101.50  155.12  1.06      1464  ! E12520 DNA encoding an antigen
gb_pat:E12535           +  101.50  155.12  1.06      1464  ! E12535 Chlamydia pneumoniae 53
gb_pat:E12541           +  101.50  155.12  1.06      1464  ! E12541 DNA encoding an antigen
gb_pat:E12525           +  101.50  152.55  1.48      1947  ! E12525 DNA encoding a fusion p
gb_pat:E12538           +  101.50  152.55  1.48      1947  ! E12538 DNA encoding a fusion p
gb_pat:A27258           +   90.50  148.59  2.46       414  ! A27258 H.sapiens TGR-CLIbis. 10
gb_pl2:AF145590         -  100.50  148.55  2.47      2537  ! AF145590 Triticum aestivum put
gb_ba1:APRTXIII         +  102.00  146.79  3.10      4048  ! X68815 A.pleuropneumoniae gene
gb_pl1:TAGL1AX1         -   97.50  142.52  5.36      2885  ! X61009 Wheat Glu-1Ax1 gene for
gb_pl1:TAU19774         -   97.50  141.97  5.75      3067  ! U19774 Triticum aestivum high-
gb_pl1:NFENOL           +   94.00  140.26  7.15      1968  ! X80474 N.frontalis enolase gen
gb_pl1:CSNADP           -   94.00  139.49  7.89      2145  ! X58832 C.sorokiniana NADP-gdh
gb_ba2:AE001328         +  100.50  135.65  12.92     10653  ! AE001328 Chlamydia trachomati
gb_ba1:AVINIFN          +   90.00  135.36  13.40     1647  ! M15815 Azotobacter vinelandii
gb_ro:RNU47331          -   94.50  135.00  14.03     3868  ! U47331 Rattus norvegicus metab
gb_pat:AR007269         +  108.50  134.70  14.59    50341  ! AR007269 Sequence 1 from pate
gb_in2:TNAF000606       -   92.50  134.59  14.80     2821  ! AF000606 Trichoplusia ni insec
gb_ba1:APXIIICAB        +  102.00  134.58  14.82    15746  ! X80055 A.pleuropneumoniae apx
gb_ph:MLCGA             +  108.50  134.36  15.25    52297  ! Z18946 Mycobacteriophage L5 c
gb_pat:I32211           -   95.00  134.23  15.50     4614  ! I32211 Sequence 1 from patent
gb_pl1:WHTGLUTB         -   92.50  134.04  15.88     2999  ! M22209 Tricticum aestivum, hig
gb_ba1:YEPYOPBDX        +   92.50  133.72  16.54     3106  ! L06215 Yersinia pseudotubercul
gb_ro:RATMGLUR4A        -   94.50  133.67  16.66     4488  ! M90518 Rat metabotropic glutam
gb_ba2:AE001662         -  101.50  133.13  17.85    16903  ! AE001662 Chlamydia pneumoniae
gb_pl1:WHTGLUTA         -   93.00  132.64  19.01     3836  ! M22208 Tricticum aestivum high
gb_pl1:SCFL01           -   95.00  132.26  19.94     5743  ! X78160 S.cerevisiae flocculati
gb_pl1:TAGL1D1B         -   92.00  131.61  21.69     3590  ! X12928 Wheat Glu-1D-1d gene fo
gb_pl1:TAGLB11B         -   92.50  131.37  22.35     4034  ! X13927 Wheat Glu-B1-1b gene fo
gb_ba1:ACFIXLJ          +   90.00  131.12  23.09     2640  ! X56658 A.caulinodans fixL and
gb_ba1:MLURHO           -   89.00  130.34  25.51     2402  ! L27277 Micrococcus luteus Rho
gb_pl2:S80654           -   84.50  130.15  26.16     1088  ! S80654 major allergen Pha a
gb_pat:A31057           -   84.50  128.95  30.51     1243  ! A31057 L.perenne mRNA 12R f
gb_ba2:AF111267         -   87.50  128.92  30.63     2146  ! AF111267 Streptomyces cattl

seq_name: gb_ba1:SDU66877

seq_documentation_block:
LOCUS       SDU66877     2995 bp    DNA             BCT       29-AUG-1996
DEFINITION  Salmonella dublin SicA (sicA) and SipB (sipB) genes, complete cds.
ACCESSION   U66877
NID         g1513311
VERSION     U66877.1  GI:1513311
KEYWORDS    .
SOURCE      Salmonella dublin.
  ORGANISM  Salmonella dublin
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Salmonella.
REFERENCE   1  (bases 1 to 2995)
  AUTHORS   Mullan,P.B., Gautier,A.V., Wood,M.W., Edwards,M.H., Jones,B.V. and
            Galyov,E.E.
  TITLE     Sip invasins of Salmonella dublin are involved in macrophage
            cytotoxicity and mouse virulence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2995)
  AUTHORS   Mullan,P.B., Gautier,A.V., Wood,M.W., Edwards,M.H., Jones,B.V. and
            Galyov,E.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-AUG-1996) Institute for Animal Health, Compton,
            Newbury, Berkshire, RG20 7NN, UK
FEATURES             Location/Qualifiers
     source          1. .2995
                     /organism="Salmonella dublin"
                     /strain="2229"
                     /db_xref="taxon:593"
     gene            131. .628
                     /gene="sicA"
     CDS             131. .628
                     /gene="sicA"
                     /codon_start=1
                     /transl_table=11
                     /product="SicA"
                     /protein_id="AAB06794.1"
                     /db_xref="PID:g1513312"
                     /db_xref="GI:1513312"
                     /translation="MDYQNNVSEERVAEMIWDAVSEGATLKDVHGIPQDMMDGLYAHA
                     YEFYNQGRLDEAETFFRFLCIYDFYNPDYTMGLAAVCQLKKQFQKACDLYAVAFTLLK
                     NDYRPVFFTGQCQLLMRKAAKARQCFELVNERTEDESLRAKALVYLEALKTAETEQHS
                     EQEKE"
     gene            631. .2412
                     /gene="sipB"
     CDS             631. .2412
                     /gene="sipB"
                     /function="virulence-associated secreted protein"
                     /codon_start=1
                     /transl_table=11
                     /product="SipB"
                     /protein_id="AAB06795.1"
                     /db_xref="PID:g1513313"
                     /db_xref="GI:1513313"
                     /translation="MVNDASSISRSGYTQNPRLAEAAFEGVRKNTDFLKAADKAFKDV
                     VATKAGDLKAGTKSGESAINTVGLKPPTDAARERLSSEGQLTLLLGKLMTLLGDVSIS
                     QLESRLAVWQAMIESQKEMGIQVSKEFQTALGEAQEATDLYEASIKKTDTAKSVYDAA
                     TKKLTQAQNKLQSLDPADPGYAQAEAAVEQAGKEATEAKEALDKATDATVKAGTDAKA
                     KAEKADNILTKFQGTANAASQNQVSQGEQDNLSNVARLTMLMAMFIEIVGKNTEESLQ
                     NDLALFNALQEGRQAEMEKKSAEFQEETRKAEETNRIMGCIGKVLGALLTIVSVVAAV
                     FTGGASLALAAVGLAVMVADEIVKAATGVSFIQQALNPIMEHVLKPLMELIGKAITKA
                     LEGLGVDKKTAEMAGSIVGAIVAAIAMVAVIVVVAVVGKGAAAKLGNALSKMMGETIK
                     KLVPNVLKQLAQNGSKLFTQGMQRITSGLGNVGSKMGLQTNALSKELVGNTLNKVALG
                     MEVTNTAAQSAGGVAEGVFIKNASEALADFMLARFAMDQIQQWLKQSVEIFGENQKVT
                     AELQKAMSSAVQQNADASRFILRQSRA"
BASE COUNT      851 a    614 c    847 g    683 t
ORIGIN
```

```
alignment_scores:
            Quality: 1202.50              Length:    292
              Ratio:  4.175                Gaps:      4
 Percent Similarity:  98.630   Percent Identity: 98.288

alignment_block:
 US-09-068-804-5 x SDU66877   ..

 Align seg 1/1  to: SDU66877  from: 1  to: 2995

    1 ArgLysAlaGluGluThrAsnArgIleMetGlyCysIleGlyLysValLe 17
      ||||||||||||||||||||||||||||||||||||||||||||||||||||
 1543 CGCAAAGCCGAGGAAACGAACCGCATTATGGGATGTATCGGGAAAGTCCT 1592

   17 uGlyAlaLeuLeuThrIleValSerValValAlaAlaValPheThrGlyG 34
      ||||||||||||||||||||||||||||||||||||||||||||||||||||
 1593 CGGCGCGCTGCTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTG 1642

   34 lyAlaSerLeuAlaLeuAlaAlaValGlyLeuAlaValMetValAlaAsp 50
      ||||||||||||||||||||||||||||||||||||||||||||||||||||
 1643 GGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTGGCCGAT 1692

   51 GluIleValLysAlaAlaThrGlyValSerPheIleGlnGlnAlaLeuAs 67
      ||||||||||||||||||||||||||||||||||||||||||||||||||||
 1693 GAAATTGTGAAGGCGGCGACGGGGGTGTCGTTTATTCAGCAGGCGCTAAA 1742

   67 nProIleMetGluHisValLeuLysProLeuMetGluLeuIleGlyLysA 84
      ||||||||||||||||||||||||||||||||||||||||||||||||||||
 1743 CCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTGATTGGCAAGG 1792

   84 laIleThrLysAlaLeuGluGlyLeuGlyValAsp.ArgLysArgGlnAr 100
      |||||||||||||||||||||||||||||||| ||||||||||||||||||
 1793 CGATTACCAAAGCGCTGGAAGGGTTAGGCGTCGATAAGAAAACGGCAGAG 1842

  100 gTrpProAlaAlaLeuLeuValArgLeuSerProLeu.CysHisGlyAsp 116
      |||||||||||||||||||||||||||||||| |||||||||
 1843 ATGGCCGGCAGCATTGTTGGTGCGATTGTCGCCGCTATTGCCATGGT... 1889

  117 Ala.ValIleValValValAlaValValGlyLysGlyAlaAlaAlaLysL 133
      ::: ||||||||||||||||||||||||||||||||||||||||||||||||
 1890 GGCGGTCATTGTGGTGGTCGCAGTTGTCGGGAAAGGCGCGGCGGCGAAAC 1939

  133 euGlyAsnAlaLeuSerLysMetMetGlyGluThrIleLysLysLeuVal 149
      ||||||||||||||||||||||||||||||||||||||||||||||||||||
 1940 TGGGTAACGCGCTGAGCAAAATGATGGGCGAAACGATTAAGAAGTTGGTG 1989

  150 ProAsnValLeuLysGlnLeuAlaGlnAsnGlySerLysLeuPheThrGl 166
      ||||||||||||||||||||||||||||||||||||||||||||||||||||
 1990 CCTAACGTGCTGAAACAGTTGGCGCAAAACGGCAGCAAACTCTTTACCCA 2039

  166 nGlyMetGlnArgIleThrSerGlyLeuGlyAsnValGlySerLysMetG 183
      ||||||||||||||||||||||||||||||||||||||||||||||||||||
 2040 GGGGATGCAACGTATTACTAGCGGTCTGGGTAATGTAGGTAGCAAGATGG 2089

  183 lyLeuGlnThrAsnAlaLeuSerLysGluLeuValGlyAsnThrLeuAsn 199
      ||||||||||||||||||||||||||||||||||||||||||||||||||||
 2090 GCCTGCAAACGAATGCCTTAAGTAAAGAGCTGGTAGGTAATACCCTAAAT 2139

  200 LysValAlaLeuGlyMetGluValThrAsnThrAlaAlaGlnSerAlaGl 216
      ||||||||||||||||||||||||||||||||||||||||||||||||||||
 2140 AAAGTGGCGTTGGGCATGGAAGTCACGAATACCGCAGCCCAGTCAGCCGG 2189

  216 yGlyValAlaGluGlyValPheIleLysAsnAlaSerGluAlaLeuAlaA 233
      ||||||||||||||||||||||||||||||||||||||||||||||||||||
 2190 TGGTGTTGCCGAGGGCGTATTTATTAAAAATGCCAGCGAGGCGCTTGCTG 2239

  233 spPheMetLeuAlaArgPheAlaMetAspGlnIleGlnGlnTrpLeuLys 249
      ||||||||||||||||||||||||||||||||||||||||||||||||||||
 2240 ATTTTATGCTCGCCCGTTTTGCCATGGATCAGATTCAGCAGTGGCTTAAA 2289

  250 GlnSerValGluIlePheGlyGluAsnGlnLysValThrAlaGluLeuGl 266
      ||||||||||||||||||||||||||||||||||||||||||||||||||||

 2290 CAATCCGTAGAAATATTTGGTGAAAACCAGAAGGTAACGGCGGAACTGCA 2339

  266 nLysAlaMetSerSerAlaValGlnGlnAsnAlaAspAlaSerArgPheI 283
      ||||||||||||||||||||||||||||||||||||||||||||||||||||
 2340 AAAAGCCATGTCTTCTGCGGTACAGCAAAATGCGGATGCTTCGCGTTTTA 2389

  283 leLeuArgGlnSerArgAla 289
      |||||||||||||||||||
 2390 TTCTGCGCCAGAGTCGCGCA 2409

seq_name: gb_ba2:STU30491

seq_documentation_block:
LOCUS       STU30491     3441 bp    DNA             BCT       19-JAN-1996
DEFINITION  Salmonella typhimurium SspA (sspA) and SspB (sspB) genes, partial
            cds, and SspC (sspC) and SspD (sspD) genes, complete cds.
ACCESSION   U30491
NID         g1117811
VERSION     U30491.1  GI:1117811
KEYWORDS    .
SOURCE      Salmonella typhimurium.
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Salmonella.
REFERENCE   1  (bases 1 to 3441)
  AUTHORS   Hueck,C.J., Hantman,M.J., Bajaj,V., Johnston,C., Lee,C.A. and
            Miller,S.I.
  TITLE     Salmonella typhimurium secreted invasion determinants are
            homologous to Shigella Ipa proteins
  JOURNAL   Mol. Microbiol. 18 (3), 479-490 (1995)
  MEDLINE   96342378
REFERENCE   2  (bases 1 to 3441)
  AUTHORS   Miller,S.I.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUN-1995) Samuel I. Miller, Infectious Disease Unit,
            Massachusetts General Hospital, Harvard Medical School, Fruit
            Street, Boston, MA 02114, USA
FEATURES             Location/Qualifiers
     source          1. .3441
                     /organism="Salmonella typhimurium"
                     /strain="SL1344"
                     /db_xref="taxon:602"
                     /map="59 min"
     gene            1. .870
                     /gene="sspB"
     CDS             <1. .870
                     /gene="sspB"
                     /note="secreted protein"
                     /codon_start=1
                     /transl_table=11
                     /function="invasion determinant"
                     /product="SspB"
                     /protein_id="AAC43546.1"
                     /db_xref="PID:g1117812"
                     /db_xref="GI:1117812"
                     /translation="RKAEETNRIMGCIGKVLGALLTIVSVVAAVFTGGASLALAAVGL
                     AVMVADEIVKAATGVSFIQQALNPIMEHVLKPLMELIGKAITKALEGLGVDKKTAEMA
                     GSIVGAIVAAIAMVAVIVVVAVVGKGAAAKLGNALSKMMGETIKKLVPNVLKQLAQNG
                     SKLFTQGMQRITSGLGNVGSKMGLQTNALSKELVGNTLNKVALGMEVTNTAAQSAGGV
                     AEGVFIKNASEALADFMLARFAMDQIQQWLKQSVEIFGENQKVTAELQKAMSSAVQQN
                     ADASRFILRQSRA"
     gene            898. .2127
                     /gene="sspC"
     CDS             898. .2127
                     /gene="sspC"
                     /note="secreted protein"
                     /codon_start=1
                     /transl_table=11
                     /function="invasion determinant"
                     /evidence=experimental
                     /product="SspC"
                     /protein_id="AAC43547.1"
                     /db_xref="PID:g1117813"
```

```
                     /db_xref="GI:1117813"
                     /translation="MLISNVGINPAAYLNNHSVENSSQTASQSVSAKDILNSIGISSS
                     KVSDLGLSPTLSAPAPGVLTQTPGTITSFLKASIQNTDMNQDLNALANNVTTKANEVV
                     QTQLREQQAEVGKFFDISGMSSSAVALLAAANTLMLTLNQADSKLSGKLSLVSFDAAK
                     TTASSMMREGMNALSGSISQSALQLGITGVGAKLEYKGLQNERGALKHNAAKIDKLTT
                     ESHSIKNVLNGQNSVKLGAEGVDSLKSLNMKKTGTDATKNLNDATLKSNAGTSATESL
                     GIKDSNKQISPEHQAILSKRLESVESDIRLEQNTMDMTRIDARKMQMTGDLIMKNSVT
                     VGGIAGASGQYAATQERSEQQISQVNNRVASTASDEARESSRKSTSLIQEMLKTMESI
                     NQSKASALAAIAGNIRA"
     gene            2198. .3229
                     /gene="sspD"
     CDS             2198. .3229
                     /gene="sspD"
                     /note="secreted protein"
                     /codon_start=1
                     /transl_table=11
                     /function="invasion determinant"
                     /evidence=experimental
                     /product="SspD"
                     /protein_id="AAC43548.1"
                     /db_xref="PID:g1117814"
                     /db_xref="GI:1117814"
                     /translation="MLNIQNYSASPHPGIVAERPQTPSASEHVETAVVPSTTEHRGTD
                     IISLSQAATKIHQAQQTLQSTPPISEENNDERTLARQQLTSSLNALAKSGVSLSAEQN
                     ENLRSAFSAPTSALFSASPMAQPRTTISDAEIWDMVSQNISAIGDSYLGVYENVVAVY
                     TDFYQAFSDILSKMGGWLLPGKDGNTVKLDVTSLKNDLNSLVNKYNQINSNTVLFPAQ
                     SGSGVKVATEAEARQWLSELNLPNSCLKSYGSGYVVTVDLTPLQKMVQDIDGLGAPGK
                     DSKLEMDNAKYQAWQSGFKAQEENMKTTLQTLTQKYSNANSLYDNLVKVLSSTISSSL
                     ETAKSFLQG"
     gene            3248. .3441
                     /gene="sspA"
     CDS             3248. .>3441
                     /gene="sspA"
                     /note="secreted protein"
                     /codon_start=1
                     /transl_table=11
                     /function="unknown"
                     /product="SspA"
                     /protein_id="AAC43549.1"
                     /db_xref="PID:g1117815"
                     /db_xref="GI:1117815"
                     /translation="MVTSVRTQPPVIMPGMQTEIKTQATNLAANLSAVRESATATLSG
                     EIKGPQLEDFPALIKQASLDA"
BASE COUNT      978 a    759 c    925 g    779 t
ORIGIN


alignment_scores:
            Quality: 1202.50             Length:    292
              Ratio:  4.175                Gaps:      4
 Percent Similarity:  98.630   Percent Identity: 98.288

alignment_block:
 US-09-068-804-5 x STU30491   ..

 Align seg 1/1  to: STU30491  from: 1  to: 3441

       1 ArgLysAlaGluGluThrAsnArgIleMetGlyCysIleGlyLysValLe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 CGCAAAGCCGAGGAAACGAACCGCATTATGGGATGTATCGGGAAAGTCCT 50

      17 uGlyAlaLeuLeuThrIleValSerValValAlaAlaValPheThrGlyG 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 CGGCGCGCTGCTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTG 100

      34 lyAlaSerLeuAlaLeuAlaAlaValGlyLeuAlaValMetValAlaAsp 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 GGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTGGCCGAT 150

      51 GluIleValLysAlaAlaThrGlyValSerPheIleGlnGlnAlaLeuAs 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 GAAATTGTGAAGGCGGCGACGGGAGTGTCGTTTATTCAGCAGGCGCTAAA 200

      67 nProIleMetGluHisValLeuLysProLeuMetGluLeuIleGlyLysA 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 CCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTGATTGGCAAGG 250

      84 laIleThrLysAlaLeuGluGlyLeuGlyValAsp.ArgLysArgGlnAr 100
         |||||||||||||||||||||||||||||||||| ||||||||||||||
     251 CGATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAG 300

     100 gTrpProAlaAlaLeuLeuValArgLeuSerProLeu.CysHisGlyAsp 116
         |||||||||||||||||||||||||||||||| |||||||||
     301 ATGGCCGGCAGCATTGTTGGTGCGATTGTCGCCGCTATTGCCATGGT... 347

     117 Ala.ValIleValValValAlaValValGlyLysGlyAlaAlaAlaLysL 133
         ::: ||||||||||||||||||||||||||||||||||||||||||||||
     348 GGCGGTCATTGTGGTGGTCGCAGTTGTCGGGAAAGGCGCGGCGGCGAAAC 397

     133 euGlyAsnAlaLeuSerLysMetMetGlyGluThrIleLysLysLeuVal 149
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     398 TGGGTAACGCGCTGAGCAAAATGATGGGCGAAACGATTAAGAAGTTGGTG 447

     150 ProAsnValLeuLysGlnLeuAlaGlnAsnGlySerLysLeuPheThrGl 166
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     448 CCTAACGTGCTGAAACAGTTGGCGCAAAACGGCAGCAAACTCTTTACCCA 497

     166 nGlyMetGlnArgIleThrSerGlyLeuGlyAsnValGlySerLysMetG 183
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     498 GGGGATGCAACGTATTACTAGCGGTCTGGGTAATGTGGGTAGCAAGATGG 547

     183 lyLeuGlnThrAsnAlaLeuSerLysGluLeuValGlyAsnThrLeuAsn 199
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     548 GCCTGCAAACGAATGCCTTAAGTAAAGAGCTGGTAGGTAATACCCTAAAT 597

     200 LysValAlaLeuGlyMetGluValThrAsnThrAlaAlaGlnSerAlaGl 216
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     598 AAAGTGGCGTTGGGCATGGAAGTCACGAATACCGCAGCCCAGTCAGCCGG 647

     216 yGlyValAlaGluGlyValPheIleLysAsnAlaSerGluAlaLeuAlaA 233
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     648 TGGTGTTGCCGAGGGCGTATTTATTAAAAATGCCAGCGAGGCGCTTGCTG 697

     233 spPheMetLeuAlaArgPheAlaMetAspGlnIleGlnGlnTrpLeuLys 249
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     698 ATTTTATGCTCGCCCGTTTTGCCATGGATCAGATTCAGCAGTGGCTTAAA 747

     250 GlnSerValGluIlePheGlyGluAsnGlnLysValThrAlaGluLeuGl 266
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     748 CAATCCGTAGAAATATTTGGTGAAAACCAGAAGGTAACGGCGGAACTGCA 797

     266 nLysAlaMetSerSerAlaValGlnGlnAsnAlaAspAlaSerArgPheI 283
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     798 AAAAGCCATGTCTTCTGCGGTACAGCAAAATGCGGATGCTTCGCGTTTTA 847

     283 leLeuArgGlnSerArgAla 289
         |||||||||||||||||||
     848 TTCTGCGCCAGAGTCGCGCA 867

seq_name: gb_ba1:STU25631

seq_documentation_block:
LOCUS       STU25631     3622 bp    DNA             BCT       06-SEP-1995
DEFINITION  Salmonella typhimurium invasin homologs SicA (sicA), SipB (sipB)
            and SipC  (sipC) genes, complete cds.
ACCESSION   U25631
NID         g975293
VERSION     U25631.1  GI:975293
KEYWORDS    .
SOURCE      Salmonella typhimurium.
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Salmonella.
REFERENCE   1  (bases 1 to 3622)
  AUTHORS   Kaniga,K., Tucker,S., Trollinger,D. and Galan,J.E.
  TITLE     Homologs of the Shigella IpaB and IpaC invasins are required for
```

```
           Salmonella typhimurium entry into cultured epithelial cells
  JOURNAL  J. Bacteriol. 177 (14), 3965-3971 (1995)
  MEDLINE  95332200
REFERENCE  2  (bases 1 to 3622)
  AUTHORS  Galan,J.E.
    TITLE  Direct Submission
  JOURNAL  Submitted (23-APR-1995) Jorge E. Galan, Molecular Genetics and
           Microbiology, School of Medicine, SUNY Stony Brook, Stony Brook, NY
           11794-5222, USA
FEATURES             Location/Qualifiers
     source          1. .3622
                     /organism="Salmonella typhimurium"
                     /strain="SL1344"
                     /db_xref="taxon:602"
     gene            75. .572
                     /gene="sicA"
     CDS             75. .572
                     /gene="sicA"
                     /codon_start=1
                     /transl_table=11
                     /product="SicA"
                     /protein_id="AAA75168.1"
                     /db_xref="PID:g975294"
                     /db_xref="GI:975294"
                     /translation="MDYQNNVSEERVAEMIWDAVSEGATLKDVHGIPQDMMDGLYAHA
                     YEFYNQGRLDEAETFFRFLCIYDFYNPDYTMGLAAVCQLKKQFQKACDLYAVAFTLLK
                     NDYRPVFFTGQCQLLMRKAAKARQCFELVNERTEDESLRAKALVYLEALKTAETEQHS
                     EQEKE"
     gene            575. .2356
                     /gene="sipB"
     CDS             575. .2356
                     /gene="sipB"
                     /codon_start=1
                     /transl_table=11
                     /product="SipB"
                     /protein_id="AAA75169.1"
                     /db_xref="PID:g975295"
                     /db_xref="GI:975295"
                     /translation="MVNDASSISRSGYTQNPRLAEAAFEGVRKNTDFLKAADKAFKDV
                     VATKAGDLKAGTKSGESAINTVGLKPPTDAAREKLSSEGQLTLLLGKLMTLLGDVSLS
                     QLESRLAVWQAMIESQKEMGIQVSKEFQTALGEAQEATDLYEASIKKTDTAKSVYDAA
                     TKKLTQAQNKLQSLDPADPGYAQAEAAVEQAGKEATEAKEALDKATDATVKAGTDAKA
                     KAEKADNILTKFQGTANAASQNQVSQGEQDNLSNVARLTMLMAMFIEIVGKNTEESLQ
                     NDLALPNALQEGRQAEMEKKSAEFQEETRKAEETNRIMGCIGKVLGALLTIVSVVAAV
                     FTGGASLALAAVGLAVMVADEIVKAATGVSFIQQALNPIMEHVLKPLMELIGKAITKA
                     LEGLGVDKKTAEMAGSIVGAIVAAIAMVAVIVVVAVVGKGAAAKLGNALSKMMGETIK
                     KLVPNVLKQLAQNGSKLFTQGMQRITSGLGNVGSKMGLQTNALSKELVGNTLNKVALG
                     MEVTNTAAQSAGGVAEGVFIKNASEALADFMLARFAMDQIQQWLKQSVEIFGENQKVT
                     AELQKAMSSAVQQNADASRFILRQSRA"
     gene            2384. .3613
                     /gene="sipC"
     CDS             2384. .3613
                     /gene="sipC"
                     /codon_start=1
                     /transl_table=11
                     /product="SipC"
                     /protein_id="AAA75170.1"
                     /db_xref="PID:g975296"
                     /db_xref="GI:975296"
                     /translation="MLISNVGINPAAYLNNHSVENSSQTASQSVSAKDILNSIGISSS
                     KVSDLGLSPTLSAPAPGVLTQTPGTITSSLKASIQNTDMNQDLNALANNVTTKANEVV
                     QTQLREQQAEVGKFFDISGMSSSAVALLAAANTLMLTLNQADSKLSGKLSLVSFDAAK
                     TTASSMMREGMNALSGSISQSALQLGITGVGAKLEYKGLQNERGALKHNAAKIDKLTT
                     ESHSIKNVLNGQNSVKLGAEGVDSLKSLNIRKPVPMRRKILMMRRLNLMPEPAPRKVW
                     VLKTVINKVSLNIYILSKRLESVESDIRLEQNYMDITRIDSAQDADDGRSDYEELGHG
                     RWYCRGVRAVRRYSGNVSEQQTSQVNNRVASTASDEARESSRKSTSLIQEMLKTMESI
                     NQSKASALAAIAGNIRA"
BASE COUNT     1049 a    759 c   1008 g    806 t
ORIGIN


alignment_scores:
          Quality: 1202.50            Length:    292
```

```
            Ratio:  4.175              Gaps:     4
 Percent Similarity:  98.630   Percent Identity: 98.288

alignment_block:
 US-09-068-804-5 x STU25631   ..

 Align seg 1/1  to: STU25631  from: 1  to: 3622

      1 ArgLysAlaGluGluThrAsnArgIleMetGlyCysIleGlyLysValLe 17
        |||||||||||||||||||||||||||||||||||||||||||||||||||
   1487 CGCAAAGCCGAGGAAACGAACCGCATTATGGGATGTATCGGGAAAGTCCT 1536

     17 uGlyAlaLeuLeuThrIleValSerValValAlaAlaValPheThrGlyG 34
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1537 CGGCGCGCTGCTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTG 1586

     34 lyAlaSerLeuAlaLeuAlaAlaValGlyLeuAlaValMetValAlaAsp 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1587 GGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTGGCCGAT 1636

     51 GluIleValLysAlaAlaThrGlyValSerPheIleGlnGlnAlaLeuAs 67
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1637 GAAATTGTGAAGGCGGCGACGGGAGTGTCGTTTATTCAGCAGGCGCTAAA 1686

     67 nProIleMetGluHisValLeuLysProLeuMetGluLeuIleGlyLysA 84
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1687 CCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTGATTGGCAAGG 1736

     84 laIleThrLysAlaLeuGluGlyLeuGlyValAsp.ArgLysArgGlnAr 100
        |||||||||||||||||||||||||||||||||| ||||||||||||||
   1737 CGATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAG 1786

    100 gTrpProAlaAlaLeuLeuValArgLeuSerProLeu.CysHisGlyAsp 116
        |||||||||||||||||||||||||||||||||| |||||||||
   1787 ATGGCCGGCAGCATTGTTGGTGCGATTGTCGCCGCTATTGCCATGGT... 1833

    117 Ala.ValIleValValValAlaValValGlyLysGlyAlaAlaAlaLysL 133
        ::: ||||||||||||||||||||||||||||||||||||||||||||||
   1834 GGCGGTCATTGTGGTGGTCGCAGTTGTCGGGAAAGGCGCGGCGGCGAAAC 1883

    133 euGlyAsnAlaLeuSerLysMetMetGlyGluThrIleLysLysLeuVal 149
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1884 TGGGTAACGCGCTGAGCAAAATGATGGGCGAAACGATTAAGAAGTTGGTG 1933

    150 ProAsnValLeuLysGlnLeuAlaGlnAsnGlySerLysLeuPheThrGl 166
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1934 CCTAACGTGCTGAAACAGTTGGCGCAAAACGGCAGCAAACTCTTTACCCA 1983

    166 nGlyMetGlnArgIleThrSerGlyLeuGlyAsnValGlySerLysMetG 183
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1984 GGGGATGCAACGTATTACTAGCGGTCTGGGTAATGTGGGTAGCAAGATGG 2033

    183 lyLeuGlnThrAsnAlaLeuSerLysGluLeuValGlyAsnThrLeuAsn 199
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   2034 GCCTGCAAACGAATGCCTTAAGTAAAGAGCTGGTAGGTAATACCCTAAAT 2083

    200 LysValAlaLeuGlyMetGluValThrAsnThrAlaAlaGlnSerAlaGl 216
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   2084 AAAGTGGCGTTGGGCATGGAAGTCACGAATACCGCAGCCCAGTCAGCCGG 2133

    216 yGlyValAlaGluGlyValPheIleLysAsnAlaSerGluAlaLeuAlaA 233
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   2134 TGGTGTTGCCGAGGGCGTATTTATTAAAAATGCCAGCGAGGCGCTTGCTG 2183

    233 spPheMetLeuAlaArgPheAlaMetAspGlnIleGlnGlnTrpLeuLys 249
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   2184 ATTTTATGCTCGCCCGTTTTGCCATGGATCAGATTCAGCAGTGGCTTAAA 2233

    250 GlnSerValGluIlePheGlyGluAsnGlnLysValThrAlaGluLeuGl 266
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   2234 CAATCCGTAGAAATATTTGGTGAAAACCAGAAGGTAACGGCGGAACTGCA 2283

    266 nLysAlaMetSerSerAlaValGlnGlnAsnAlaAspAlaSerArgPheI 283
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||
      2284 AAAAGCCATGTCTTCTGCGGTACAGCAAAATGCGGATGCTTCGCGTTTTA 2333

       283 leLeuArgGlnSerArgAla 289
           |||||||||||||||||||
      2334 TTCTGCGCCAGAGTCGCGCA 2353

seq_name: gb_ba1:STSIPGNS

seq_documentation_block:
LOCUS       STSIPGNS     5393 bp    DNA             BCT       04-OCT-1995
DEFINITION  S.typhi spaT, sipB, sipC, sipD and sipA genes.
ACCESSION   X82670
NID         g1009001
VERSION     X82670.1  GI:1009001
KEYWORDS    sipA gene; sipB gene; sipC gene; sipD gene; spaT gene.
SOURCE      Salmonella typhi.
  ORGANISM  Salmonella typhi
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Salmonella.
REFERENCE   1  (bases 1 to 5393)
  AUTHORS   Hermant,D., Menard,R., Arricau,N., Parsot,C. and Popoff,M.Y.
  TITLE     Functional conservation of the Salmonella and Shigella effectors of
            entry into epithelial cells
  JOURNAL   Mol. Microbiol. 17 (4), 781-789 (1995)
  MEDLINE   96111497
REFERENCE   2  (bases 1 to 5393)
  AUTHORS   Popoff,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-1994) M. Popoff, Institut Pasteur, Unite des
            Enterobacteries, 28 rue du Docteur Roux, 75724 Paris Cedex 15,
            FRANCE
FEATURES             Location/Qualifiers
     source          1. .5393
                     /organism="Salmonella typhi"
                     /strain="Ty2"
                     /db_xref="taxon:601"
     gene            43. .540
                     /gene="spaT"
     CDS             43. .540
                     /gene="spaT"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA57987.1"
                     /db_xref="PID:g1009002"
                     /db_xref="GI:1009002"
                     /db_xref="SPTREMBL:Q57260"
                     /translation="MDYQNNVSEERVAEMIWDAVSEGATLKDVHGIPQDMMDGLYAHA
                     YEFYNQGRLDEAETFFRFLCIYDFYNPDYTMGLAAVCQLKKQFQKACDLYAVAFTLLK
                     NDYRPVFFTGQCQLLMRKAAKARQCFELVNERTEDESLRAKALVYLEALKTAETEQHS
                     EQEKE"
     gene            543. .2324
                     /gene="sipB"
     CDS             543. .2324
                     /gene="sipB"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA57988.1"
                     /db_xref="PID:g1009003"
                     /db_xref="GI:1009003"
                     /db_xref="SPTREMBL:Q56134"
                     /translation="MVNDASSISRSGYTQNPRLAEAAFEGVRKNTDFLKAADKAFKDV
                     VATKAGDLKAGTKSGESAINTVGLKPPTDAAREKLSSEGQLTLLLGKLMTLLGDVSLS
                     QLESRLAVWQAMIESQKEMGIQVSKEFQTALGEAQEATDLYEASIKKTDTAKSVYDAA
                     AKKLTQAQNKLQSLDPADPGYAQAEAAVEQAGKEATEAKEALDKATDATVKAGTDAKA
                     KAEKADNILTKFQGTANAASQNQVSQGEQDNLSNVARLTMLMAMFIEIVGKNTEESLQ
                     NDLALFNALQEGRQAEMEKKSAEFQEETRKAEETNRIMGCIGKVLGALLTIVSVVAAV
                     FTGGASLALAAVGLAVMVADEIVKAATGVSFIQQALNPIMEHVLKPLMELIGKAITKA
                     LEGLGVDKKTAEMAGSIVGAIVAAIAMVAVIVVVAVVGKGAAAKLGNALSKMMGETIK
                     KLVPNVLKQLAQNGSKLFTQGMQRITSGLGNVGSKMGLQTNALSKELVGNTLNKVALG
                     MEVTNTAAQSAGGVAEGVFIKNASEALADFMLARFAMDQIQQWLKQSVEIFGENQKVT
                     AELQKAMSSAVQQNADASRFILRQSRA"
     gene            2352. .3581
                     /gene="sipC"
     CDS             2352. .3581
                     /gene="sipC"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA57989.1"
                     /db_xref="PID:g1009004"
                     /db_xref="GI:1009004"
                     /db_xref="SPTREMBL:Q56135"
                     /translation="MLISNVGINPAAYLNNHSVENSSQTASQSVSAKDILNSIGISSS
                     KVSDLGLSPTLSAPAPGVLTQTPGTITSFLKASIQNTDMNQDLNALANNVTTKANEVV
                     QTQLREQQAEVGKFFDISGMSSSAVALLAAANTLMLTLNQADSKLSGKLSLVSFDAAK
                     TTASSMMREGMNALSGSISQSALQLGITGVGAKLEYKGLQNERGALKHNAAKIDKLTT
                     ESHSIKNVLNGQNSVKLGAEGVDSLKSLNMKKTGTDATKNLNDATLKSNAGTSATESL
                     GIKNSNKQISPEHQAILSKRLESVESDIRLEQNTMDMTRIDARKMQMTGDLIMKNSVT
                     VGGIAGASRQYAATQERSEQQISQVNNRVASTASDEARESSRKSTSLIQEMLKTMESI
                     NQSKASALAAIAGNIRA"
     gene            3651. .4670
                     /gene="sipD"
     CDS             3651. .4670
                     /gene="sipD"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA57990.1"
                     /db_xref="PID:g1009005"
                     /db_xref="GI:1009005"
                     /db_xref="SPTREMBL:Q56136"
                     /translation="MLNIQNYSASPHPGIVAERPQTPSASEHAEIAVVPSTTEHRGTD
                     IISLSQAATKIQQAQQTLQSTPPISEENNDERTLARQQLTSSLNALAKSGVSLSAEQN
                     ENLRSTFSARRRPYLALRLWPARTTISDAEIWDMVSQNISAIGDSYLGVYENVVAVYT
                     DFYQAFSDILSKMGGWLSPGKDGNTIKLNVDSLKSEISSLINKYTQINKNTILFPSQI
                     GSGMTTATKAEAEQWIKELNLPDSCLKASGSGYVVLVDTGPLSKMVSDLNGIGSGSAL
                     ELDNAKYQAWQSGFKAQEENLKTTLQILTQKYSNANSLYDNLVKVLSSTISSSLETAK
                     SFLQG"
     gene            4689. .5393
                     /gene="sipA"
     CDS             4689. .>5393
                     /gene="sipA"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA57991.1"
                     /db_xref="PID:g1009006"
                     /db_xref="GI:1009006"
                     /db_xref="SPTREMBL:Q56137"
                     /translation="MVTSVRTQPPVIMPGMQTEIKTQATNLAANLSAVRESATATLSG
                     EIKGQQLEDFPALIKQASLDALFKCGKDAEALKEVFTNSNNVAGKKAIMEFAGLFRSA
                     LNATSDSPEAKTLLMKVGAEYTAQIIKDGLKEKSAFGPWLPETKKAEAKLENLEKQLL
                     DIIKNNTGGELSKLSTNLVMQEVMPYIASCIEHNFGCTLDPLTRSSLTQLVDKAAAKA
                     VEALDMCHQKLTQEQGT"
BASE COUNT     1584 a    1156 c    1458 g    1195 t
ORIGIN


alignment_scores:
           Quality: 1196.50              Length:    292
             Ratio:  4.155                 Gaps:      4
 Percent Similarity:  98.630   Percent Identity: 97.945

alignment_block:
 US-09-068-804-5 x STSIPGNS   ..

 Align seg 1/1  to: STSIPGNS  from: 1  to: 5393

       1 ArgLysAlaGluGluThrAsnArgIleMetGlyCysIleGlyLysValLe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1455 CGCAAAGCCGAGGAAACGAACCGCATTATGGGATGTATCGGGAAAGTCCT 1504

      17 uGlyAlaLeuLeuThrIleValSerValValAlaAlaValPheThrGlyG 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1505 CGGCGCGCTGCTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTG 1554

      34 lyAlaSerLeuAlaLeuAlaAlaValGlyLeuAlaValMetValAlaAsp 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
1555 GGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTGGCCGAT 1604

  51 GluIleValLysAlaAlaThrGlyValSerPheIleGlnGlnAlaLeuAs 67
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1605 GAAATTGTGAAGGCGGCGACGGGGGTGTCGTTTATTCAGCAGGCGCTAAA 1654

  67 nProIleMetGluHisValLeuLysProLeuMetGluLeuIleGlyLysA 84
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1655 CCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTGATTGGCAAGG 1704

  84 laIleThrLysAlaLeuGluGlyLeuGlyValAsp.ArgLysArgGlnAr 100
     |||||||||||||||||||||||||||||||| ||||||||||||||||
1705 CGATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAG 1754

 100 gTrpProAlaAlaLeuLeuValArgLeuSerProLeu.CysHisGlyAsp 116
     ||||:::|||||||||||||||||||||||||||| |||||||||
1755 ATGGCAGGCAGCATTGTTGGTGCGATTGTCGCCGCTATTGCCATGGT... 1801

 117 Ala.ValIleValValValAlaValValGlyLysGlyAlaAlaAlaLysL 133
     ::: ||||||||||||||||||||||||||||||||||||||||||||||
1802 AGCGGTCATTGTGGTGGTCGCAGTTGTCGGGAAAGGCGCGGCGGCGAAAC 1851

 133 euGlyAsnAlaLeuSerLysMetMetGlyGluThrIleLysLysLeuVal 149
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1852 TGGGTAACGCGCTGAGCAAAATGATGGGCGAAACGATTAAGAAGTTGGTG 1901

 150 ProAsnValLeuLysGlnLeuAlaGlnAsnGlySerLysLeuPheThrGl 166
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1902 CCTAACGTGCTGAAACAGTTGGCACAAAACGGCAGCAAACTCTTTACCCA 1951

 166 nGlyMetGlnArgIleThrSerGlyLeuGlyAsnValGlySerLysMetG 183
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1952 GGGGATGCAACGTATTACTAGCGGCCTGGGTAATGTGGGTAGCAAGATGG 2001

 183 lyLeuGlnThrAsnAlaLeuSerLysGluLeuValGlyAsnThrLeuAsn 199
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2002 GCCTGCAAACGAATGCCTTAAGTAAAGAGCTGGTAGGTAATACCCTAAAT 2051

 200 LysValAlaLeuGlyMetGluValThrAsnThrAlaAlaGlnSerAlaGl 216
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2052 AAAGTGGCGTTGGGCATGGAAGTCACGAATACCGCAGCCCAGTCAGCCGG 2101

 216 yGlyValAlaGluGlyValPheIleLysAsnAlaSerGluAlaLeuAlaA 233
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2102 TGGGGTTGCCGAGGGGGTATTTATTAAAAATGCCAGCGAGGCGCTTGCTG 2151

 233 spPheMetLeuAlaArgPheAlaMetAspGlnIleGlnGlnTrpLeuLys 249
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2152 ATTTTATGCTCGCCCGTTTTGCCATGGATCAGATTCAGCAGTGGCTTAAA 2201

 250 GlnSerValGluIlePheGlyGluAsnGlnLysValThrAlaGluLeuGl 266
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2202 CAATCCGTAGAAATATTTGGTGAAAACCAGAAGGTAACGGCGGAACTGCA 2251

 266 nLysAlaMetSerSerAlaValGlnGlnAsnAlaAspAlaSerArgPheI 283
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2252 AAAAGCCATGTCTTCTGCGGTACAGCAAAATGCGGATGCTTCGCGTTTTA 2301

 283 leLeuArgGlnSerArgAla 289
     |||||||||||||||||||
2302 TTCTGCGCCAGAGTCGCGCA 2321
```

seq_name: gb_ba1:W10PPIBCA

seq_documentation_block:

```
LOCUS       W10PPIBCA    4761 bp    DNA             BCT       07-MAY-1996
DEFINITION  Plasmid pWR100 (from S.flexneri) antigen ippI, ipaB, ipaC genes,
            complete cds, and ippD gene, 3' end.
ACCESSION   M34849
NID         g155327
VERSION     M34849.1  GI:155327
KEYWORDS    antigen.
SOURCE      Plasmid pWR100 (clone: HB101/pHS[4108,5357].).
  ORGANISM  Plasmid pWR100
            plasmids.
REFERENCE   1  (bases 1 to 4761; 1 to 538; 3987 to 4761)
  AUTHORS   Baudry,B., Kaczorek,M. and Sansonetti,P.J.
  TITLE     Nucleotide sequence of the invasion plasmid antigen B and C genes
            (ipaB and ipaC) of Shigella flexneri
  JOURNAL   Microb. Pathog. 4 (5), 345-357 (1988)
  MEDLINE   89200844
COMMENT     Draft entry and computer-readable sequence for [1] kindly submitted
            by B.Baudry, 05-JUN-1990.
               Center for Vaccine Development
               University of Maryland at Baltimore
               10 South Pine Street
               Baltimore, MD 21201.
FEATURES             Location/Qualifiers
     source          1. .4761
                     /organism="Plasmid pWR100"
                     /plasmid="Plasmid pWR100"
                     /specific_host="M90T"
                     /db_xref="taxon:2661"
                     /clone="HB101/pHS[4108,5357]."
     CDS             <1. .549
                     /note="ORF"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA98422.1"
                     /db_xref="PID:g155328"
                     /db_xref="GI:155328"
                     /translation="KLVKKILSVFNLKQRFPQKNFGCPVNINKIRDSVIDKIKDSNSG
                     NQLFCWMSQERTTYVSSMINRSIDEMAIHNGVVLTSDNKKNIFAAIEKKFPDIKLDEK
                     SAQTSISHTALNEIASSGLRAKILKRYSSDMDLFNTQMKDLTNLVSSSVYDKIFNEST
                     KVLQIEISAEVLKAVYRQSNTN"
     -10_signal      556. .561
                     /gene="ippI"
     gene            556. .1073
                     /gene="ippI"
     RBS             595. .600
                     /gene="ippI"
     CDS             606. .1073
                     /gene="ippI"
                     /codon_start=1
                     /transl_table=11
                     /product="antigen"
                     /protein_id="AAA98423.1"
                     /db_xref="PID:g155329"
                     /db_xref="GI:155329"
                     /translation="MSLNITENESISTAVIDAINSGATLKDINAIPDDMMDDIYSYAY
                     DFYNKGRIEEAEVFFRFLCIYDFYNVDYIMGLAAIYQIKEQFQQAADLYAVAFALGKN
                     DYTPVFHTGQCQLRLKAPLKAKECFELVIQHSNDEKLKIKAQSYLDAIQDIKE"
     gene            1065. .2821
                     /gene="ipaB"
     RBS             1065. .1070
                     /gene="ipaB"
     CDS             1079. .2821
                     /gene="ipaB"
                     /codon_start=1
                     /transl_table=11
                     /product="antigen"
                     /protein_id="AAA98424.1"
                     /db_xref="PID:g155330"
                     /db_xref="GI:155330"
                     /translation="MHNVSTTTTGFPLAKILTSTELGDNTIQAANDAANKLFSLTIAD
                     LTANQNINTTNAHSTSNILIPELKAPKSLNASSQLTLLIGNLIQILGEKSLTALTNKI
                     TAWKSQQQARQQKNLEFSDKINTLLSETEGLTRDYEKQINKLKNADSKIKDLENKINQ
                     IQTRLSELDPESPEKKKLSREEIQLTIKKDAAVKDRTLIEQKTLSIHSKLTDKSMQLE
                     KEIDSFSAFSNTASAEQLSTQQKSLTGLASVTQLMATFIQLVGKNNEESLKNDLALFQ
                     SLQESRKTEMERKSDEYAAEVRKAEELNRVMGCVGKILGALLTIVSVVAAAFSGGASL
                     ALAAVGLALMVTDAIVQAATGNSFMEQALNPIMKAVIEPLIKLLSDAFTKMLEGLGVD
                     SKKAKMIGSILGAIAGALVLVAAVVLVATVGKQAAAKLAENIGKIIGKTLTDLIPKFL
                     KNFSSQLDDLITNAVARLNKFLGAAGDEVISKQIISTHLNQAVLLGESVNSATQAGGS
                     VASAVFQNSASTNLADLTLSKYQVEQLSKYISEAIEKFGQLQEVIADLLASMSNSQAN
                     RTDVAKAILQQTTA"
     gene            2723. .3932
```

```
                /gene="ipaC"
    -35_signal  2723. .2728
                /gene="ipaC"
    -10_signal  2745. .2750
                /gene="ipaC"
    RBS         2829. .2834
                /gene="ipaC"
    CDS         2841. .3932
                /gene="ipaC"
                /codon_start=1
                /transl_table=11
                /product="antigen"
                /protein_id="AAA98425.1"
                /db_xref="PID:g155331"
                /db_xref="GI:155331"
                /translation="MEIQNTKPTQTLYTDISTKQTQSSSETQKSQNYQQIAAHIPLNV
                GKNPVLTTTLNDDQLLKLSEQVQHDSEIIARLTDKKMKDLSEMSHTLTPENTLDISSL
                SSNAVSLIISVAVLLSALRTAETKLGSQLSLIAFDATKSAAENIVRQGLAALSSSITG
                AVTQVGITGIGAKKTHSGISDQKGALRKNLATAQSLEKELAGSKLGLNKQIDTNITSP
                QTNSSTKFLGKNKLAPDNISLSTEHKTSLSSPDISLQDKIDTQRRTYELNTLSAQQKQ
                NIGRATMETSAVAGNISTSGGRYASALEEEEQLISQASSKQAEEASQVSKEASQATNQ
                LIQKLLNIIDSINQSKNSAASQIAGNIRA"
    RBS         3971. .3977
                /gene="ipaD"
    gene        3971. .4761
                /gene="ipaD"
    CDS         3983. .>4761
                /gene="ipaD"
                /codon_start=1
                /transl_table=11
                /product="antigen"
                /protein_id="AAA98426.1"
                /db_xref="PID:g552045"
                /db_xref="GI:552045"
                /translation="MNITTLTNSISTSSFSPNNTNGSSTETVNSDIKTTTSSHPVSSL
                TMLNDTLHNIRTTNQALKKELSQKTLTKTSLEEIALHSSQISMDVNKSAQLLDILSRN
                EYPINKDARELLHSAPKEAELDGDQMISHRELWAKIANSINDINEQYLKVYEHAVSSY
                TQMYQDFSAVLSSLAGWISPGGNDGNSVKLQVNSLKKALEELKEKYKDKPLYPANNTV
                SQEQANKWLTELGGTIGKVSQKNGGYVVSINMTPIDNMLKS"
BASE COUNT     1769 a    952 c    805 g   1235 t
ORIGIN


alignment_scores:
            Quality:  557.00            Length:    289
              Ratio:  2.411               Gaps:      1
 Percent Similarity:  79.931   Percent Identity: 46.021

alignment_block:
 US-09-068-804-5 x W10PPIBCA   ..

 Align seg 1/1  to: W10PPIBCA  from: 1  to: 4761

       1 ArgLysAlaGluGluThrAsnArgIleMetGlyCysIleGlyLysValLe 17
         ||||||||||||||||   ||||||:::|||||||||:::|||||:::||
   1970 CGTAAAGCAGAAGAACTCAACAGAGTAATGGGTTGTGTTGGGAAAATACT 2019

      17 uGlyAlaLeuLeuThrIleValSerValValAlaAlaValPheThrGlyG 34
         |||||||||||||||||||||||||||||||||||||:::|||:::||||
    2020 TGGGGCACTTTTAACTATCGTTAGTGTTGTTGCAGCAGCTTTTTCTGGAG 2069

      34 lyAlaSerLeuAlaLeuAlaAlaValGlyLeuAlaValMetValAlaAsp 50
         |||||||||||||||||||||||||||||||||||:::|||||:::|||
    2070 GAGCCTCTCTAGCACTGGCAGCTGTTGGTTTAGCTCTTATGGTTACGGAT 2119

      51 GluIleValLysAlaAlaThrGlyValSerPheIleGlnGlnAlaLeuAs 67
         :::||||||:::|||||||||||||   ||||||::::::||||||||||
    2120 GCTATAGTACAAGCAGCGACCGGCAATTCCTTCATGGAACAAGCCCTGAA 2169

      67 nProIleMetGluHisValLeuLysProLeuMetGluLeuIleGlyLysA 84
         ||||||||||~~  |||:::~~ ||||||:::~~|||::::::   |
    2170 TCCGATCATGAAAGCAGTCATTGAACCCTTAATCAAACTCCTTTCAGATG 2219

      84 laIleThrLysAlaLeuGluGlyLeuGlyValAspArgLysArgGlnArg 100
         ||:::||||||   ||||||||||||||||||||   |||:::   :::
    2220 CATTTACAAAAATGCTCGAAGGCTTGGGCGTCGACTCGAAAAAAGCCAAA 2269

     101 TrpProAlaAlaLeuLeuValArgLeuSerProLeuCysHisGlyAspAl 117
               ::::::::::|||       ::::::                   ||
    2270 ATGATTGGCTCTATTCTGGGGGCAATCGCAGGCGCTCTTGTCCTAGTTGC 2319

     117 aValIleValValValAlaValValGlyLysGlyAlaAlaAlaLysLeuG 134
         |:::::::|||:::|||||||:::|||||||||   |||||||||||||||:
    2320 AGCAGTCGTTCTCGTAGCCACTGTTGGTAAACAGGCAGCAGCAAAACTTG 2369

     134 lyAsnAlaLeuSerLysMetMetGlyGluThrIleLysLysLeuValPro 150
         ::::::::::::::|||:::::::||   |||::::::   |||:::|||
    2370 CAGAAAATATTGGCAAAATAATAGGTAAAACCCTCACAGACCTTATACCA 2419

     151 AsnValLeuLysGlnLeuAlaGlnAsnGlySerLysLeuPheThrGlnGl 167
         :::   ||||||:::::::::   :::   :::   |||:::|||:::::
    2420 AAGTTTCTCAAGAATTTTTCTTCTCAACTGGACGATTTAATCACTAATGC 2469

     167 yMetGlnArgIleThrSerGlyLeuGlyAsnValGlySerLysMetGlyL 184
         ::::   |||::::::      |||||||:::::|||:::
    2470 TGTTGCCAGATTAAATAAATTTCTTGGTGCAGCGGGTGAT..... 2509

     184 euGlnThrAsnAlaLeuSerLysGluLeuValGlyAsnThrLeuAsnLys 200
                 :::::::::|||||:::::::::::::::   |||||||:::
    2510 ........GAAGTAATATCCAAACAAATTATTTCCACCCATTTAAACCAA 2551

     201 ValAlaLeuGlyMetGluValThrAsnThrAlaAlaGlnSerAlaGlyGl 217
         ::::::|||      |||   :::|||:::|||:::|||::::::::||::
    2552 GCAGTTTTATTAGGAGAAAGTGTTAACTCTGCCACACAAGCGGGAGGAAG 2601

     217 yValAlaGluGlyValPheIleLysAsnAlaSerGluAlaLeuAlaAspP 234
         :||||||   :::|||||||   ::::::|||||   :::|||||||||:
    2602 TGTCGCTTCTGCTGTTTTCCAGAACAGCGCGTCGACAAATCTAGCAGACC 2651

     234 heMetLeuAlaArgPheAlaMetAspGlnIleGlnGlnTrpLeuLysGln 250
         ::   |||:::::::::   ::::::|||:::   :::   :::   :::
    2652 TGACATTATCGAAATATCAAGTTGAACAACTGTCAAAATATATCAGTGAA 2701

     251 SerValGluIlePheGlyGluAsnGlnLysValThrAlaGluLeuGlnLy 267
         :::::::|||   ||||||:::   |||   |||::::|||:::|||
    2702 GCAATAGAAAAATTCGGCCAATTGCAGGAAGTAATTGCAGATCTATTAGC 2751

     267 sAlaMetSerSerAlaValGlnGlnAsnAlaAspAlaSerArgPheIleL 284
          :::|||||||::::::      :::   :::|||:::::::::   ||||
    2752 CTCAATGTCCAACTCTCAGGCTAATAGAACTGATGTTGCAAAAGCAATTT 2801

     284 euArgGlnSerArgAla 289
         ||:::|||:::   |||
    2802 TGCAACAAACTACTGCT 2818

seq_name: gb_bal:SHFIPABCD

seq_documentation_block:
LOCUS       SHFIPABCD    5255 bp    DNA             BCT       19-MAR-1991
DEFINITION  S.flexneri ipaBCD region encoding invasion plasmid antigens.
ACCESSION   J04117
NID         g152740
VERSION     J04117.1  GI:152740
KEYWORDS    ipaA gene; ipaB gene; ipaC gene; ipaD; membrane protein.
SOURCE      S.flexeri (strain M90T-W) DNA, clone pHC17.
  ORGANISM  Shigella flexneri
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Shigella.
REFERENCE   1  (bases 1 to 5255)
  AUTHORS   Venkatesan,M.M., Buysse,J.M. and Kopecko,D.J.
  TITLE     Characterization of invasion plasmid antigen genes (ipaBCD) from
            Shigella flexneri
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 85, 9317-9321 (1988)
  MEDLINE   89057927
COMMENT     Draft entry and clean copy of sequence [1] kindly submitted by
            M.M.Venkatesan, 03-JAN-1989.
```

```
FEATURES             Location/Qualifiers
     source          1. .5255
                     /organism="Shigella flexneri"
                     /db_xref="taxon:623"
     CDS             606. .1073
                     /note="18 kD protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA26521.1"
                     /db_xref="PID:g152741"
                     /db_xref="GI:152741"
                     /translation="MSLNITENESISTAVIDAINSGATLKDINAIPDDMMDDIYSYAY
                     DFYNKGRIEEAEVFFRFLCIYDFYNVDYIMGLAAIYQIKEQFQQAADLYAVAFALGKN
                     DYTPVFHTGQCQLRLKAPLKAKECFELVIQHSNDEKLKIKAQSYLDAIQDIKE"
     CDS             1079. .2821
                     /note="IpaB protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA26522.1"
                     /db_xref="PID:g152742"
                     /db_xref="GI:152742"
                     /translation="MHNVSTTTTGFPLAKILTSTELGDNTIQAANDAANKLFSLTIAD
                     LTANQNINTTNAHSTSNILIPELKAPKSLNASSQLTLLIGNLIQILGEKSLTALTNKI
                     TAWKSQQQARQQKNLEFSDKINTLLSETEGLTRDYEKQINKLKNADSKIKDLENKINQ
                     IQTRLSNLDPESPEKKKLSREEIQLTIKKDAAVKDRTLIEQKTLSIHSKLTDKSMQLE
                     KEIDSFSAFSNTASAEQLSTQQKSLTGLASVTQLMATFIQLVGKNNEESLKNDLALFQ
                     SLQESRKTEMERKSDEYAAEVRKAEELNRVMGCVGKILGALLTIVSVVAAAFSGGASL
                     ALAAVGLALMVTDAIVQAATGNSFMEQALNPIMKAVIEPLIKLLSDAFTKMLEGLGVD
                     SKKAKMIGSILGAIAGALVLVAAVVLVATVGKQAAAKLAENIGKIIGKTLTDLIPKFL
                     KNFSSQLDDLITNAVARLNKFLGAAGDEVISKQIISTHLNQAVLLGESVNSATQAGGS
                     VASAVFQNSASTNLADLTLSKYQVEQLSKYISEAIEKFGQLQEVIADLLASMSNSQAN
                     RTDVAKAILQQTTA"
     CDS             2784. .3932
                     /note="IpaC protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA26523.1"
                     /db_xref="PID:g152743"
                     /db_xref="GI:152743"
                     /translation="MLQKQFCNKLLLDTNKENVMEIQNTKPTQTLYTDISTKQTQSSS
                     ETQKSQNYQQIAAHIPLNVGKNPVLTTTLNDDQLLKLSEQVQHDSEIIARLTDKKMKD
                     LSEMSHTLTPENTLDISSLSSNAVSLIISVAVLLSALRTAETKLGSQLSLIAFDATKS
                     AAENIVRQGLAALSSSITGAVTQVGITGIGAKKTHSGISDQKGALRKNLATAQSLEKE
                     LAGSKLGLNKQIDTNITSPQTNSSTKFLGKNKLAPDNISLSTEHKTSLSSPDISLQDK
                     IDTQRRTYELNTLSAQQKQNIGRATMETSAVAGNISTSGGRYASALEEEEQLISQASS
                     KQAEEASQVSKEASQATNQLIQKLLNIIDSINQSKNSAASQIAGNIRA"
     CDS             3983. .4981
                     /note="IpaD protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA26524.1"
                     /db_xref="PID:g152744"
                     /db_xref="GI:152744"
                     /translation="MNITTLTNSISTSSFSPNNTNGSSTETVNSDIKTTTSSHPVSSL
                     TMLNDTLHNIRTTNQALKKELSQKTLTKTSLEEIALHSSQISMDVNKSAQLLDILSRN
                     EYPINKDARELLHSAPKEAELDGDQMISHRELWAKIANSINDINEQYLKVYEHAVSSY
                     TQMYQDFSAVLSSLAGWISPGGNDGNSVKLQVNSLKKALEELKEKYKDKPLYPANNTV
                     SQEQANKWLTELGGTIGKVSQKNGGYVVSINMTPIDNMLKSLDNLGGNGEVVLDNAKY
                     QAWNAGFSAEDETMKNNLQTLVQKYSNANSIFDNLVKVLSSTISSCTDTDKLFLHF"
     CDS             4990. .>5255
                     /note="IpaA protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA26525.1"
                     /db_xref="PID:g551981"
                     /db_xref="GI:551981"
                     /translation="MHNVNNTQAPTFLYKATSPSSTEYSELKSKISDIHSSQTSLKTP
                     ASVSEKENFATSFNQKCLDFLFSSSGKEDVLRSIYSNSMNAYAK"
BASE COUNT      1950 a    1037 c     889 g    1379 t
ORIGIN

alignment_scores:
```

```
            Quality:  557.00          Length:    289
              Ratio:  2.411            Gaps:      1
Percent Similarity:  79.931   Percent Identity: 46.021

alignment_block:
 US-09-068-804-5 x SHFIPABCD   ..

 Align seg 1/1  to: SHFIPABCD  from: 1  to: 5255

       1 ArgLysAlaGluGluThrAsnArgIleMetGlyCysIleGlyLysValLe 17
         ||||||||||||||||   ||||||:::||||||||:::||||||:::||
    1970 CGTAAAGCAGAAGAACTCAACAGAGTAATGGGTTGTGTTGGGAAAATACT 2019

      17 uGlyAlaLeuLeuThrIleValSerValValAlaAlaValPheThrGlyG 34
         ||||||||||||||||||||||||||||||||||||||:::|||:::||||
    2020 TGGGGCACTTTTAACTATCGTTAGTGTTGTTGCAGCAGCTTTTTCTGGAG 2069

      34 lyAlaSerLeuAlaLeuAlaAlaValGlyLeuAlaValMetValAlaAsp 50
         ||||||||||||||||||||||||||||||||:::||||||:::|||
    2070 GAGCCTCTCTAGCACTGGCAGCTGTTGGTTTAGCTCTTATGGTTACGGAT 2119

      51 GluIleValLysAlaAlaThrGlyValSerPheIleGlnGlnAlaLeuAs 67
         :::||||||:::||||||||||||   ||||||::::::||||||||||||
    2120 GCTATAGTACAAGCAGCGACCGGCAATTCCTTCATGGAACAAGCCCTGAA 2169

      67 nProIleMetGluHisValLeuLysProLeuMetGluLeuIleGlyLysA 84
         |||||||||   |||:::  ||||||:::  |||:::::  |
    2170 TCCGATCATGAAAGCAGTCATTGAACCCTTAATCAAACTCCTTTCAGATG 2219

      84 laIleThrLysAlaLeuGluGlyLeuGlyValAspArgLysArgGlnArg 100
         ||:::|||||  ||||||||||||||||||||  |||:::  :::
    2220 CATTTACAAAAATGCTCGAAGGCTTGGGCGTCGACTCGAAAAAAGCCAAA 2269

     101 TrpProAlaAlaLeuLeuValArgLeuSerProLeuCysHisGlyAspAl 117
         :::::::::|||      ::::::                    ||
    2270 ATGATTGGCTCTATTCTGGGGGCAATCGCAGGCGCTCTTGTCCTAGTTGC 2319

     117 aValIleValValValAlaValValGlyLysGlyAlaAlaAlaLysLeuG 134
         |:::::|||:::|||||||:::|||||||  ||||||||||||||:
    2320 AGCAGTCGTTCTCGTAGCCACTGTTGGTAAACAGGCAGCAGCAAAACTTG 2369

     134 lyAsnAlaLeuSerLysMetMetGlyGluThrIleLysLysLeuValPro 150
         :::::::::::::|||:::::|||  |||::::::  |||:::|||
    2370 CAGAAAATATTGGCAAAATAATAGGTAAAACCCTCACAGACCTTATACCA 2419

     151 AsnValLeuLysGlnLeuAlaGlnAsnGlySerLysLeuPheThrGlnGl 167
         :::  ||||||:::::::  :::  :::  |||:::|||:::::
    2420 AAGTTTCTCAAGAATTTTTCTTCTCAACTGGACGATTTAATCACTAATGC 2469

     167 yMetGlnArgIleThrSerGlyLeuGlyAsnValGlySerLysMetGlyL 184
         ::::  |||::::::      ||||||:::::|||:::
    2470 TGTTGCCAGATTAAATAAATTTCTTGGTGCAGCGGGTGAT.......... 2509

     184 euGlnThrAsnAlaLeuSerLysGluLeuValGlyAsnThrLeuAsnLys 200
               ::::::::::|||||:::::::::::::::  ||||||:::
    2510 ........GAAGTAATATCCAAACAAATTATTTCCACCCATTTAAACCAA 2551

     201 ValAlaLeuGlyMetGluValThrAsnThrAlaAlaGlnSerAlaGlyGl 217
         :::::::|||      |||  :::|||:::|||:::|||:::::::|||::
    2552 GCAGTTTTATTAGGAGAAAGTGTTAACTCTGCCACACAAGCGGGAGGAAG 2601

     217 yValAlaGluGlyValPheIleLysAsnAlaSerGluAlaLeuAlaAspP 234
         :||||||   :::||||||  ::::::||||||  :::||||||||||:
    2602 TGTCGCTTCTGCTGTTTTCCAGAACAGCGCGTCGACAAATCTAGCAGACC 2651

     234 heMetLeuAlaArgPheAlaMetAspGlnIleGlnGlnTrpLeuLysGln 250
         ::  |||:::::::::   :::::::|||:::  :::  :::  :::
    2652 TGACATTATCGAAATATCAAGTTGAACAACTGTCAAAATATATCAGTGAA 2701

     251 SerValGluIlePheGlyGluAsnGlnLysValThrAlaGluLeuGlnLy 267
         :::::::|||  ||||||:::  |||  |||:::|||::|||
    2702 GCAATAGAAAAATTCGGCCAATTGCAGGAAGTAATTGCAGATCTATTAGC 2751
```

```
    267 sAlaMetSerSerAlaValGlnGlnAsnAlaAspAlaSerArgPheIleL 284
        :::||||||::::::      :::   :::|||::::::::::   ||||
   2752 CTCAATGTCCAACTCTCAGGCTAATAGAACTGATGTTGCAAAAGCAATTT 2801

    284 euArgGlnSerArgAla 289
        ||:::|||:::   |||
   2802 TGCAACAAACTACTGCT 2818

seq_name: gb_ba1:SFPVR2

seq_documentation_block:
LOCUS       SFPVR2       7663 bp    DNA             BCT       12-SEP-1993
DEFINITION  Shigella flexineri plasmid YSH6000 virulence region-2 DNA
            containing 24k gene, 18k gene, ipaB gene, ipaC gene and ipaD gene.
ACCESSION   X15319
NID         g47052
VERSION     X15319.1  GI:47052
KEYWORDS    ipaB gene; ipaC gene; ipaD gene; virulence proteins.
SOURCE      Enterobacteriaceae.
  ORGANISM  Enterobacteriaceae
            Bacteria; Proteobacteria; gamma subdivision.
REFERENCE   1  (bases 1 to 7663)
  AUTHORS   Sasakawa,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAY-1989) Sasakawa C., Institute of Medical Science,
            University of Tokyo, Shirokanedai 4-6-1, Minato-ku Tokyo 108, Japan
REFERENCE   2  (bases 1010 to 6538)
  AUTHORS   Sasakawa,C., Adler,B., Tobe,T., Okada,N., Nagai,S., Komatsu,K. and
            Yoshikawa,M.
  TITLE     Functional organization and nucleotide sequence of virulence
            Region-2 on the large virulence plasmid in Shigella flexneri 2a
  JOURNAL   Mol. Microbiol. 3 (9), 1191-1201 (1989)
  MEDLINE   90014179
COMMENT     Region 2 has three promoters that transcribe RNAs of 4.0, 4.5 and
            7.5kb. The 4.0 kb transcript produces the 24,18kD, IpaB and IpaC
            proteins. The 4.5 kb transcript produces the ipaD protein. The
            7.5kb transcript covers the entire region 2.
            Data kindly reviewed (20-NOV-1989) by Sasakawa C.
FEATURES             Location/Qualifiers
     source          1. .7663
                     /organism="Enterobacteriaceae"
                     /plasmid="YSH6000"
                     /db_xref="taxon:543"
     promoter        1291. .1296
                     /note="-35 region for 4kb transcript"
     promoter        1312. .1318
                     /note="-10 region for 4kb transcript"
     RBS             1360. .1362
                     /note="put. ribosome binding site"
     CDS             1370. .1996
                     /note="24k gene (AA 1-208)"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA33379.1"
                     /db_xref="PID:g47053"
                     /db_xref="GI:47053"
                     /db_xref="SWISS-PROT:P33548"
                     /translation="MQILNKILPQVEFAIPRPSFDSLSRNKLVKKILSVFNLKQRFPQ
                     KNFGCPVNINKIRDSVIDKIKDSNSGNQLFCWMSQERTTYVSSMINRSIDEMAIHNGV
                     VLTSDNKKNIFAAIEKKFPDIKLDEKSAQTSISHTALNEIASSGLRAKILKRYSSDMD
                     LFNTQMKDLTNLVSSSVYDKIFNESTKVLQIEISAEVLKAVYRQSNTN"
     RBS             2043. .2045
                     /note="put. ribosome binding site"
     CDS             2053. .2520
                     /note="18k gene (AA 1-155)"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA33380.1"
                     /db_xref="PID:g47054"
                     /db_xref="GI:47054"
                     /db_xref="SWISS-PROT:P18008"
                     /translation="MSLNITENESISTAVIDAINSGATLKDINAIPDDMMDDIYSYAY
                     DFYNKGRIEEAEVFFRFLCIYDFYNVDYIMGLAAIYQIKEQFQQAADLYAVAFALGKN
                     DYTPVFHTGQCQLRLKAPLKAKECFELVIQHSNDEKLKIKAQSYLDAIQDIKE"
     RBS             2513. .2515
                     /note="put. ribosome binding site"
     CDS             2526. .4268
                     /note="ipaB gene (AA 1-580)"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA33381.1"
                     /db_xref="PID:g47055"
                     /db_xref="GI:47055"
                     /db_xref="SWISS-PROT:P18011"
                     /translation="MHNVSTTTTGFPLAKILASTELGDNTIQAANDAANKLFSLTIAD
                     LTANQNINTTNAHSTSNILIPELKAPKSLNASSQLTLLIGNLIQILGEKSLTALTNKI
                     TAWKSQQQARQQKNLEFSDKINTLLSETEGLTRDYEKQINKLKNADSKIKDLENKINQ
                     IQTRLSELDPESPEKKKLSREEIQLTIKKDAAVKDRTLIEQKTLSIHSKLTDKSMQLE
                     KEIDSFSAFSNTASAEQLSTQQKSLTGLASVTQLMATFIQLVGKNNEESLKNDLALFQ
                     SLQESRKTEMERKSDEYAAEVRKAEELNRVMGCVGKILGALLTIVSVVAAAFSGGASL
                     ALAAVGLALMVTDAIVQAATGNSFMEQALNPIMKAVIEPLIKLLSDAFTKMLEGLGVD
                     SKKAKMIGSILGAIAGALVLVAAVVLVATVGKQAAAKLAENIGKIIGKTLTDLIPKFL
                     KNFSSQLDDLITNAVARLNKFLGAAGDEVISKQIISTHLNQAVLLGESVNSATQAGGS
                     VASAVFQNSASTNLADLTLSKYQVEQLSKYISEAIEKFGQLQEVIADLLASMSNSQAN
                     RTDVAKAILQQTTA"
     RBS             4224. .4226
                     /note="put. ribosome binding site"
     CDS             4231. .5379
                     /note="ipaC gene (AA 1-382)"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA33382.1"
                     /db_xref="PID:g47056"
                     /db_xref="GI:47056"
                     /db_xref="SWISS-PROT:P18012"
                     /translation="MLQKQFCNKLLLDTNKENVMEIQNTKPTQILYTDISTKQTQSSS
                     ETQKSQNYQQIAAHIPLNVGKNPVLTITTLNDDQLLKLSEQVQHDSEIIARLTDKKMKD
                     LSEMSHTLTPENTLDISSLSSNAVSLIISVAVLLSALRTAETKLGSQLSLIAFDATKS
                     AAENIVRQGLAALSSSITGAVTQVGITGIGAKKTHSGISDQKGALRKNLATAQSLEKE
                     LAGSKLGLNKQIDTNITSPQTNSSTKFLGKNKLAPDNISLSTEHKTSLSSPDISLQDK
                     IDTQRRTYELNTLSAQQKQNIGRATMETSAVAGNISTSGGRYASALEEEEQLISQASS
                     KQAEEASQVSKEASQATNQLIQKLLNIIDSINQSKNSTASQIAGNIRA"
     promoter        5376. .5381
                     /note="-35 region for 4.5kb transcript"
     promoter        5396. .5401
                     /note="-10 region for 4.5kb transcript"
     RBS             5421. .5423
                     /note="put. ribosome binding site"
     CDS             5430. .6428
                     /note="ipaD gene (AA 1-332)"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA33383.1"
                     /db_xref="PID:g47057"
                     /db_xref="GI:47057"
                     /db_xref="SWISS-PROT:P18013"
                     /translation="MNITTLTNSISTSSFSPNNTNGSSTETVNSDIKTTTSSHPVSSL
                     TMLNDTLHNIRTTNQALKKELSQKTLTKTSLEEIALHSSQISMDVNKSAQLLDILSRH
                     EYPINKDARELLHSAPKEAELDGDQMISHRELWAKIANSINDINEQYLKVYEHAVSSY
                     TQMYQDFSAVLSSLAGWISPGGNDGNSVKLQVNSLKKALEELKEKYKDKPLYPANNTV
                     SQEQANKWLTELGGTIGKVSQKNGGYVVSINMTPIDNMLKSLDNLGGNGEVVLDNAKY
                     QAWNAGFSAEDETMKNNLQTLVQKYSNANSIFDNLVKVLSSTISSCTDTDKLFLHF"
BASE COUNT     2845 a   1501 c   1246 g   2071 t
ORIGIN


alignment_scores:
            Quality:  557.00              Length:    289
              Ratio:  2.411                 Gaps:      1
 Percent Similarity:  79.931   Percent Identity: 46.021

alignment_block:
 US-09-068-804-5 x SFPVR2   ..

 Align seg 1/1  to: SFPVR2  from: 1  to: 7663
```

```
     1 ArgLysAlaGluGluThrAsnArgIleMetGlyCysIleGlyLysValLe 17
       ||||||||||||||||   ||||||:::||||||||:::||||||:::||
  3417 CGTAAAGCAGAAGAACTCAACAGAGTAATGGGTTGTGTTGGGAAAATACT 3466

    17 uGlyAlaLeuLeuThrIleValSerValValAlaAlaValPheThrGlyG 34
       ||||||||||||||||||||||||||||||||||||:::|||:::||||
  3467 TGGGGCACTTTTAACTATCGTTAGTGTTGTTGCAGCAGCTTTTTCTGGAG 3516

    34 lyAlaSerLeuAlaLeuAlaAlaValGlyLeuAlaValMetValAlaAsp 50
       ||||||||||||||||||||||||||||||||||:::||||||:::|||
  3517 GAGCCTCTCTAGCACTGGCAGCTGTTGGTTTAGCTCTTATGGTTACGGAT 3566

    51 GluIleValLysAlaAlaThrGlyValSerPheIleGlnGlnAlaLeuAs 67
       :::||||||:::||||||||||||   ||||||::::::||||||||||
  3567 GCTATAGTACAAGCAGCGACCGGCAATTCCTTCATGGAACAAGCCCTGAA 3616

    67 nProIleMetGluHisValLeuLysProLeuMetGluLeuIleGlyLysA 84
       |||||||||      |||:::   ||||||:::   |||::::::   |
  3617 TCCGATCATGAAAGCAGTCATTGAACCCTTAATCAAACTCCTTTCAGATG 3666

    84 laIleThrLysAlaLeuGluGlyLeuGlyValAspArgLysArgGlnArg 100
       ||:::||||||   ||||||||||||||||||||   |||:::   :::
  3667 CATTTACAAAAATGCTCGAAGGCTTGGGCGTCGACTCGAAAAAAGCCAAA 3716

   101 TrpProAlaAlaLeuLeuValArgLeuSerProLeuCysHisGlyAspAl 117
             :::::::::|||      ::::::                  ||
  3717 ATGATTGGCTCTATTCTGGGGGCAATCGCAGGCGCTCTTGTCCTAGTTGC 3766

   117 aValIleValValValAlaValValGlyLysGlyAlaAlaAlaLysLeuG 134
       |:::::|||:::||||||:::|||||||||   ||||||||||||||:
  3767 AGCAGTCGTTCTCGTAGCCACTGTTGGTAAACAGGCAGCAGCAAAACTTG 3816

   134 lyAsnAlaLeuSerLysMetMetGlyGluThrIleLysLysLeuValPro 150
       ::::::::::::::|||::::::|||   |||::::::   |||:::|||
  3817 CAGAAAATATTGGCAAAATAATAGGTAAAACCCTCACAGACCTTATACCA 3866

   151 AsnValLeuLysGlnLeuAlaGlnAsnGlySerLysLeuPheThrGlnGl 167
       :::   ||||||:::::::::   :::   :::   |||:::|||:::::
  3867 AAGTTTCTCAAGAATTTTTCTTCTCAACTGGACGATTTAATCACTAATGC 3916

   167 yMetGlnArgIleThrSerGlyLeuGlyAsnValGlySerLysMetGlyL 184
       ::::   |||::::::      ||||||:::::::|||:::
  3917 TGTTGCCAGATTAAATAAATTTCTTGGTGCAGCGGGTGAT.......... 3956

   184 euGlnThrAsnAlaLeuSerLysGluLeuValGlyAsnThrLeuAsnLys 200
             :::::::::||||||:::::::::::::::   ||||||:::
  3957 ........GAAGTAATATCCAAACAAATTATTTCCACCCATTTAAACCAA 3998

   201 ValAlaLeuGlyMetGluValThrAsnThrAlaAlaGlnSerAlaGlyGl 217
       ::::::|||      |||   :::|||:::|||:::|||:::::::|||::
  3999 GCAGTTTATTAGGAGAAAGTGTTAACTCTGCCACACAAGCGGGAGGAAG 4048

   217 yValAlaGluGlyValPheIleLysAsnAlaSerGluAlaLeuAlaAspP 234
       :||||||   :::||||||   ::::::||||||   ::::||||||||:
  4049 TGTCGCTTCTGCTGTTTTCCAGAACAGCGCGTCGACAAATCTAGCAGACC 4098

   234 heMetLeuAlaArgPheAlaMetAspGlnIleGlnGlnTrpLeuLysGln 250
       ::   |||:::::::::   ::::::|||:::   :::   :::   :::
  4099 TGACATTATCGAAATATCAAGTTGAACAACTGTCAAAATATATCAGTGAA 4148

   251 SerValGluIlePheGlyGluAsnGlnLysValThrAlaGluLeuGlnLy 267
       ::::::|||   ||||||:::   |||   |||:::|||::|||
  4149 GCAATAGAAAAATTCGGCCAATTGCAGGAAGTAATTGCAGATCTATTAGC 4198

   267 sAlaMetSerSerAlaValGlnGlnAsnAlaAspAlaSerArgPheIleL 284
       :::||||||::::::      :::   :::|||:::::::::   ||||
  4199 CTCAATGTCCAACTCTCAGGCTAATAGAACTGATGTTGCAAAAGCAATTT 4248

   284 euArgGlnSerArgAla 289
       ||:::|||:::   |||
  4249 TGCAACAAACTACTGCT 4265

seq_name: gb_ba1:SDIPABCD
```

```
seq_documentation_block:
LOCUS       SDIPABCD     6869 bp    DNA             BCT       24-JUL-1991
DEFINITION  Shigella dysenteriae ipaBCD genes.
ACCESSION   X60777
NID         g46932
VERSION     X60777.1  GI:46932
KEYWORDS    ipaB gene; ipaC gene; ipaD gene; ippI gene.
SOURCE      Shigella dysenteriae.
  ORGANISM  Shigella dysenteriae
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Shigella.
REFERENCE   1  (bases 1 to 6869)
  AUTHORS   Palchaudhuri,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUN-1991) S. Palchaudhuri, Wayne State Univers.School
            of Medicine, Department of Immunology and, Microbiology, 540 E.
            Canfield, Detroit, MI 48201, USA
REFERENCE   2  (bases 1 to 6869)
  AUTHORS   Yao,R. and Palchaudhuri,S.
  TITLE     Nucleotide sequence of the ipaBCD structure genes of Shigella
            dysenteriae
  JOURNAL   Unpublished
FEATURES             Location/Qualifiers
     source          1. .6869
                     /organism="Shigella dysenteriae"
                     /strain="CG097"
                     /db_xref="taxon:622"
     CDS             317. .943
                     /note="24kD ORF"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA43188.1"
                     /db_xref="PID:g46933"
                     /db_xref="GI:46933"
                     /db_xref="SWISS-PROT:Q04314"
                     /translation="MQILNKILPQVEFAIPRPSFNSLSRNKLVKKILSVFNLKQRFPQ
                     KNFGCTVNINKIRDNVIDKIKDSNSGNQLFCWMSQERTTYVSSMINRSIDEMAIHNGV
                     VLTSDNKKNIFAAIEKKFPDIKLDEKSAQTSISHTALNEIASSGLRAKILKRYSSDMD
                     LFNTQMKDLTNLVSSSVYDKIFNESTKVLQIEISAEVLKAVYRQSNTN"
     CDS             1000. .1467
                     /note="IppI ORF"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA43189.1"
                     /db_xref="PID:g46934"
                     /db_xref="GI:46934"
                     /db_xref="SWISS-PROT:P18008"
                     /translation="MSLNITENESISTAVIDAINSGATLKDINAIPDDMMDDIYSYAY
                     DFYNKGRIEEAEVFFRFLCIYDFYNVDYIMGLAAIYQIKEQFQQAADLYAVAFALGKN
                     DYTPVFHTGQCQLRLKAPLKAKECFELVIQHSNDEKLKIKAQSYLDAIQDIKE"
     CDS             1473. .3215
                     /note="IpaB ORF"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA43190.1"
                     /db_xref="PID:g46935"
                     /db_xref="GI:46935"
                     /db_xref="SWISS-PROT:Q03945"
                     /translation="MHNVNTTTTGLSLAKILASTELGDNTIQAGNDAANKLFSLTIAD
                     LTANKNINTTNAHSTSNILIPELKAPKSLNASSQLTLLIGNLIQILGEKSLTALTNKI
                     TAWKSQQQARQQKNLEFSDKINTLLSETEGLTRDYEKQINKLKNADSKIKDLENKINQ
                     IQTRLSELDPDSPEKKKLSREEIQLTIKKDAAVKDRTLIEQKTLSIHSKLIDKSMQLE
                     KEIDSFSAFSNTASAEQLSTQQKSLTGLASVTQLMATFIQLVGKNNEESLKNDLALFQ
                     SLQESRKTEMERKSDEYAAEVRKAEELNRVMGCVGKILGALLTIVSVVAAAFSGGASL
                     ALADVGLALMVTDAIVQAATGNSFMEQALNPIMKAVIEPLIKLLSDAFTKMLEGLGVD
                     SKKAKMIGSILGAIAGALVLVAAVVLVATVGKQAAAKLAENIGKIIGKTLTDLIPKFL
                     KNFSSQLDDLITNAVARLNKFLGAAGDEVISKQIISTHLNQAVLLGESVNSATQAGGS
                     VASAVFQNSASTNLADLTLSKYQVEQLSKYISEAIEKFGQLQEVIADLLASMSNSQAN
                     RTDVAKAILQQTTA"
     CDS             3178. .4326
                     /note="IpaC ORF"
                     /codon_start=1
```

```
                     /transl_table=11
                     /protein_id="CAA43191.1"
                     /db_xref="PID:g46936"
                     /db_xref="GI:46936"
                     /db_xref="SWISS-PROT:Q03946"
                     /translation="MLQKQFCNKLLLDTNKENVMEIQNTKSTQILYTDISTKQTQSSS
                     ETQKSQNYQQIAAHIPLNVGKNPVLTTTLNDDQLLKLSEQVQHDSEIIARLTDKKMKD
                     LSEMSHTLTPENTLDISSLSSNAVSLIISVAVLLSALRTAETKLGSQLSLIAFDATKS
                     AAENIVRQGLAALSSSITGAVTQVGITGIGAKKTHSGISDQKGALRKNLATAQSLEKE
                     LAGSKLGLNKQIDTNITSPQTNSSTKFLGKNKLAPDNISLSTEHKTSLSSPDISLQDK
                     IDTQRRTYELNTLSAQQKQNIGRATMETSAVAGNISTSGGRYASALEEEEQLISQASS
                     KQAEEASQVSKEASQATNQLIQKLLNIIDNINQSRSSTASQIAGNIRA"
     CDS             4377. .5375
                     /note="IpaD ORF"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA43192.1"
                     /db_xref="PID:g46937"
                     /db_xref="GI:46937"
                     /db_xref="SWISS-PROT:Q03947"
                     /translation="MNITTLTNSISTSSFSPNNTNGSSTETVNSDIKTTTSSHPVSSL
                     TMLNDTLHNIRTTNQALKKELSQKTLTKTSLEEIALHSSQISMDVNKSAQLLNILSKT
                     EYPINKDARELLHSAPKEAELDGYEMISHRELWAKIANSINDINEQYLKVYEHAVSSY
                     TQMYQEFSAVLSSLAGWISPGGNDGNSVKLQVKSLKDALTTLKKNYEDKPLYPATNTV
                     SEQEANKWLTELGGTIGIVSAKNGGYVVSINMTPIYNMLNRLDNLGGNGEVVLDNAKY
                     QAWNAGFSAEDETMKNNLQTLVQKYSNANSIFDNLVKVLSSTISSCTDTDKLFLHF"
BASE COUNT     2518 a   1369 c   1135 g   1847 t
ORIGIN


alignment_scores:
            Quality:  554.00            Length:    289
              Ratio:  2.398              Gaps:      1
 Percent Similarity:  79.931   Percent Identity: 45.675

alignment_block:
 US-09-068-804-5 x SDIPABCD   ..

 Align seg 1/1  to: SDIPABCD  from: 1  to: 6869

       1 ArgLysAlaGluGluThrAsnArgIleMetGlyCysIleGlyLysValLe 17
         ||||||||||||||||   ||||||:::||||||||||:::||||||:::||
    2364 CGTAAAGCAGAAGAACTCAACAGAGTAATGGGTTGTGTTGGGAAAATACT 2413

      17 uGlyAlaLeuLeuThrIleValSerValValAlaAlaValPheThrGlyG 34
         |||||||||||||||||||||||||||||||||||||:::|||:::||||
    2414 TGGGGCACTTTTAACTATCGTTAGTGTTGTTGCAGCAGCTTTTTCTGGAG 2463

      34 lyAlaSerLeuAlaLeuAlaAlaValGlyLeuAlaValMetValAlaAsp 50
         |||||||||||||||||||:::|||||||||||||:::||||||:::|||
    2464 GAGCCTCTCTAGCACTGGCAGATGTTGGTTTAGCTCTTATGGTTACGGAT 2513

      51 GluIleValLysAlaAlaThrGlyValSerPheIleGlnGlnAlaLeuAs 67
         :::||||||:::||||||||||||   ||||||::::::||||||||||||
    2514 GCTATAGTACAAGCAGCGACCGGCAATTCCTTCATGGAACAAGCCCTGAA 2563

      67 nProIleMetGluHisValLeuLysProLeuMetGluLeuIleGlyLysA 84
         ||||||||||      |||:::   ||||||:::   |||:::::   |
    2564 TCCGATCATGAAAGCAGTCATTGAACCCTTAATCAAACTCCTTTCAGATG 2613

      84 laIleThrLysAlaLeuGluGlyLeuGlyValAspArgLysArgGlnArg 100
         ||:::||||||   ||||||||||||||||||||   |||:::   :::
    2614 CATTTACAAAAATGCTCGAAGGCTTGGGCGTCGACTCGAAAAAAGCCAAA 2663

     101 TrpProAlaAlaLeuLeuValArgLeuSerProLeuCysHisGlyAspAl 117
               ::::::::::|||      ::::::                 ||
    2664 ATGATTGGCTCTATTCTGGGGGCAATCGCAGGCGCTCTTGTCCTGGTTGC 2713

     117 aValIleValValValAlaValValGlyLysGlyAlaAlaAlaLysLeuG 134
         |:::::|||:::|||||:::||||||||   ||||||||||||||||:
    2714 AGCAGTCGTTCTCGTAGCCACTGTTGGTAAACAGGCAGCAGCAAAACTTG 2763

     134 lyAsnAlaLeuSerLysMetMetGlyGluThrIleLysLysLeuValPro 150
         ::::::::::::::|||:::::|||   |||::::::   |||:::|||
    2764 CAGAGAATATTGGCAAAATAATAGGTAAAACCCTCACAGACCTTATACCA 2813

     151 AsnValLeuLysGlnLeuAlaGlnAsnGlySerLysLeuPheThrGlnGl 167
         :::   |||||:::::::::   :::   :::   |||:::|||:::::
    2814 AAGTTTCTCAAGAATTTTTCTTCTCAACTGGACGATTTAATCACTAATGC 2863

     167 yMetGlnArgIleThrSerGlyLeuGlyAsnValGlySerLysMetGlyL 184
         ::::   |||::::::      |||||:::::::|||:::
    2864 TGTTGCCAGATTAAATAAATTTCTTGGTGCAGCGGGTGAT..... 2903

     184 euGlnThrAsnAlaLeuSerLysGluLeuValGlyAsnThrLeuAsnLys 200
                ::::::::::|||||:::::::::::::::   |||||||:::
    2904 ........GAAGTAATATCCAAACAAATTATTTCCACCCATTTAAACCAA 2945

     201 ValAlaLeuGlyMetGluValThrAsnThrAlaAlaGlnSerAlaGlyGl 217
         :::::::|||      |||   :::|||:::|||:::|||:::::::|||::
    2946 GCAGTTTTATTAGGAGAAAGTGTTAACTCTGCCACACAAGCGGGAGGAAG 2995

     217 yValAlaGluGlyValPheIleLysAsnAlaSerGluAlaLeuAlaAspP 234
         :||||||   :::||||||   ::::::||||||   :::||||||||||:
    2996 TGTCGCTTCTGCTGTTTTCCAGAACAGCGCGTCGACAAATCTAGCAGACC 3045

     234 heMetLeuAlaArgPheAlaMetAspGlnIleGlnGlnTrpLeuLysGln 250
         ::   |||:::::::::   :::::::|||:::   :::   :::   :::
    3046 TGACATTATCGAAATATCAAGTTGAACAACTGTCAAAATATATCAGTGAA 3095

     251 SerValGluIlePheGlyGluAsnGlnLysValThrAlaGluLeuGlnLy 267
         :::::::|||   ||||||:::   |||   |||:::|||:::|||
    3096 GCAATAGAAAAATTCGGCCAATTGCAGGAAGTAATTGCAGATCTATTAGC 3145

     267 sAlaMetSerSerAlaValGlnGlnAsnAlaAspAlaSerArgPheIleL 284
          :::||||||::::::      :::   :::|||:::::::   ||||
    3146 CTCAATGTCCAACTCTCAGGCTAATAGAACTGATGTTGCAAAAGCAATTT 3195

     284 euArgGlnSerArgAla 289
         ||:::|||:::   |||
    3196 TGCAACAAACTACTGCT 3212

seq_name: gb_pr3:AF061832

seq_documentation_block:
LOCUS       AF061832     2453 bp    mRNA            PRI       10-MAY-1998
DEFINITION  Homo sapiens M4 protein deletion mutant mRNA, complete cds.
ACCESSION   AF061832
NID         g3126877
VERSION     AF061832.1  GI:3126877
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2453)
  AUTHORS   Zhao,Z., Huang,X., Li,N. and Cao,X.
  TITLE     A new human M4 protein with deletion
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2453)
  AUTHORS   Zhao,Z., Huang,X., Li,N. and Cao,X.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-APR-1998) Department of Immunology, Second Military
            Medical University, Shanghai Brilliance Biotechnology Institute,
            800 Xiangyin Road, Shanghai 200433, P.R. China
FEATURES             Location/Qualifiers
     source          1. .2453
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             231. .2306
                     /codon_start=1
                     /evidence=experimental
                     /product="M4 protein deletion mutant"
                     /protein_id="AAC16002.1"
                     /db_xref="PID:g3126878"
                     /db_xref="GI:3126878"
```

```
                     /translation="MAAGVEAAAEVAATEIKMEEESGAPGVPSGNGASGPKGEGERPA
                     QNEKRKEKNIKRGGNRFEPYANPTKRYRAFITNIPFDVKWQSLKDLVKEKVGEVTYVE
                     LLMDAEGKSRGCAVVEFKMEESMKKAAEVLNKHSLSGRPLKVKEDPDGECARRAMQKA
                     GRLGSTVFVANLDYKVGWKKLKEVFSMAGVVVRADILEDKDGKSRGIGTVTFEQSIEA
                     VQAISMFNGQLLFDRPMHVKMDERALPKGDFFPPERPQQLPHGLGGIGMGLGPGGQPI
                     DANHLNKGIGMGNIGPAGMGMEGIGFGINKMGGMEGPFGGGMENMGRFGSGMNMGRIN
                     EILSNALKRGEIIAKQGGGGGGGSVPGIERMGPGIDRLGGAGMERMGAGLGHGMDRVG
                     SEIERMGLVMDRMGSVERMGSGIERMGPLGLDHMASSIERMGQTMERIGSGVERMGAG
                     MGFGLERMAAPIDRVGQTIERMGSGVERMGPAIERMGPSMERMVPAGMGAGLERMGPV
                     MDRMATGLERMGANNLERMGLERMGANSLERMGLERMGANSLERMGPAMGPALGAGIE
                     RMGLAMGGGGGASFDRAIEMERGNFGGSFAGSFGGAGGHAPGVARKACQIFVRNLPFD
                     FTWKMLKDKFNECGHVLYADIKMENGKSKGCGVVKFESPEVAERACRMMNGMKLSGRE
                     IDVRIDRNA"
     misc_difference 704^705
                     /note="insertion of 117 nucleotides results in a larger
                     form of M4 protein"
BASE COUNT      601 a    528 c    839 g    485 t
ORIGIN


alignment_scores:
            Quality:  108.00              Length:    273
              Ratio:  0.831                Gaps:      6
 Percent Similarity:  47.619   Percent Identity: 17.216

alignment_block:
 US-09-068-804-5 x AF061832   ..

 Align seg 1/1  to: AF061832  from: 1  to: 2453

      10 MetGlyCysIleGlyLysValLeuGlyAlaLeuLeuThrIleValSerVa 26
         ||||||   |||:::  :::|||:::::::   |||:::        ::
    1218 ATGGGCAGGATAAATGAAATCCTAAGTAATGCACTGAAGAGAGGAGAGAT 1267

      26 lValAlaAlaValPheThrGlyGlyAlaSerLeuAlaLeuAlaAlaValG 43
         ::::|||           ||||||::::::   ::::::::::::::::
    1268 CATTGCAAAGCAGGGAGGAGGTGGAGGTGGAGGAAGCGTCCCTGGGATC. 1316

      43 lyLeuAlaValMetValAlaAspGluIleValLysAlaAlaThrGlyVal 59

    1316 ..................................................... 1316

      60 SerPheIleGlnGlnAlaLeuAsnProIleMetGluHisValLeuLysPr 76
                    :::   :::::::|||   ::::::::::::      ::
    1317 ............GAGAGGATGGGTCCTGGCATTGACCGCCTCGGGGGTGC 1354

      76 oLeuMetGluLeuIleGlyLysAlaIleThrLysAlaLeuGluGlyLeuG 93
         :   ||||||   :::|||   ::::::       :::::::::   :::|
    1355 CGGCATGGAGCGCATGGGCGCGGGCCTGGGCCACGGCATGGATCGCGTGG 1404

      93 lyValAspArgLysArgGlnArgTrpProAlaAlaLeuLeuValArgLeu 109
         ||   :::      |||
    1405 GCTCCGAGATCGAGCGC..................................... 1421

     110 SerProLeuCysHisGlyAspAlaValIleValValValAlaValValGl 126

    1421 ..................................................... 1421

     126 yLysGlyAlaAlaAlaLysLeuGlyAsnAlaLeuSerLysMet....... 140
                        :::|||   ::::::::::::|||
    1422 ..................ATGGGCCTGGTCATGGACCGCATGGGCTCCG 1452

     141 ........MetGlyGluThrIleLysLysLeuValProAsnValLeuLys 154
                 ||||||      |||   ::::::   |||       |||
    1453 TGGAGCGCATGGGCTCCGGCATTGAGCGCATGGGCCCGCTGGGCCTCGAC 1502

     155 GlnLeuAlaGlnAsnGlySerLysLeuPheThrGlnGlyMetGlnArgIl 171
         ::::::|||   :::      ::::::      |||   |||:::|||||
    1503 CACATGGCCTCCAGCATTGAGCGCATG...GGCCAGACCATGGAGCGCAT 1549

     171 eThrSerGlyLeuGlyAsnValGlySerLysMet......GlyLeuGlnT 186
         |   ||||||:::      :::|||:::   |||       ||||||:::
    1550 TGGCTCTGGCGTGGAGCGCATGGGTGCCGGCATGGGCTTCGGCCTTGAGC 1599

     186 hrAsnAlaLeuSerLysGluLeuValGlyAsnThrLeuAsnLysValAla 202
            |||   :::   :::   |||||:::|||:::::::::::::::
    1600 GCATGGCCGCTCCCATCGACCGTGTGGGCCAGACCATTGAGCGCATGGGC 1649

     203 LeuGlyMetGluValThrAsnThrAlaAlaGlnSerAlaGlyGlyValAl 219
         |||:::|||      ::::::|||   :::      |||
    1650 TCTGGCGTGGAGCGCATGGGCCCTGCCATCGAGCGCATGGGCCCGAGCAT 1699

     219 aGluGlyValPheIleLysAsnAlaSerGluAlaLeuAlaAspPheMetL 236
         |||   :::         :::   :::::::::|||:::
    1700 GGAGCGCATGGTGCCCGCAGGTATGGGAGCTGGCCTGGAG.......... 1739

     236 euAlaArgPheAlaMetAspGlnIleGlnGlnTrpLeuLysGlnServal 252
                    |||:::::::::   :::::::::      ::::::
    1740 ..........CGCATGGGCCCCGTGATGGATCGCATGGCCACCGGCCTG 1778

     253 GluIlePheGlyGluAsnGlnLysValThrAlaGluLeuGlnLysAlaMe 269
         |||      |||:::|||:::                |||::::::
    1779 GAGCGCATGGGCGCCAACAATCTGGAGCGGATGGGCCTGGAGCGCATGGG 1828

     269 tSerSerAlaValGlnGln 275
         ::::::::::::::::::
    1829 CGCCAACAGCCTCGAGCGC 1847

seq_name: gb_pri:HUMM4PRO

seq_documentation_block:
LOCUS       HUMM4PRO     2457 bp    mRNA            PRI       05-MAR-1993
DEFINITION  Human M4 protein mRNA, complete cds.
ACCESSION   L03532
NID         g187280
VERSION     L03532.1  GI:187280
KEYWORDS    M4 protein.
SOURCE      Homo sapiens (library: lambda ZAP) osteosarcoma cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2457)
  AUTHORS   Datar,K.V., Dreyfuss,G. and Swanson,M.S.
  TITLE     The Human hnRNP M Proteins: Identification of a
            Methionine/Arginine-Rich Repeat Motif in Ribonucleoproteins
  JOURNAL   Nucleic Acids Res. 21, 439-446 (1993)
  MEDLINE   93181232
FEATURES             Location/Qualifiers
     source          1. .2457
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="osteosarcoma"
                     /tissue_lib="lambda ZAP"
     CDS             12. .2201
                     /codon_start=1
                     /function="hnRNA-binding protein"
                     /product="M4 protein"
                     /protein_id="AAA36192.1"
                     /db_xref="PID:g187281"
                     /db_xref="GI:187281"
                     /translation="MAAGVEAAAEVAATEIKMEEESGGPACERQRGSGPKGEGERPAQ
                     NEKRKVKNIKRGGNRFEPYANPTKRYRAFITNIPFDVKWQSLKDLVKEKVGEVTYVEL
                     LMDAEGKSRGCAVVEFKMEESMKKAAEVLNKHSLSGRPLKVKEDPDGEHARRAMQKVM
                     ATTGGMGMGPGGPGMITIPPSILNNPNIPNEIIHALQAGRLGSTVFVANLDYKVGWKK
                     LKEVFSMAGVVVRADILEDKDGKSRGIGTVTFEQSIEAVQAISMFNGQLLFDRPMHVK
                     MDERALPKGDFFPPERPQQLPHGLGGIGMGLGPGGQPIDANHLNKGIGMGNIGPAGMG
                     MEGIGFGINKMGGMEGPFGGGMENMGRFGSGMNMGRINEILSNALKRGEIIAKQGGGG
                     GGGSVPGIERMGPGIDRLGGAGMERMGAGLGHGMDRVGSEIERMGLVMDRMGSVERMG
                     SGIERMGPLGLDHMASSIERMGQTMERIGSGVERMGAGMGFGLERMAAPIDRVGQTIE
                     RMGSGVERMGPAIERMGPSMERMVPAGMGAGLERMGPVMDRMATGLERMGANNLERMG
                     LERMGANSLERMGLERMGANSLERMGPAMGPALGAGIERMGLAMGGGGGASFDRAIEM
                     ERGNFGGSFAGSFGGAGGHAPGVARKACQIFVRNLPFDFTWKMLKDKFNECGHVLYAD
                     IKMENGKSKGCGVVKFESPEVAERACRMMNGMKLSGREIDVRIDRNA"
     polyA_signal    2440. .2445
     polyA_site      2457
BASE COUNT      623 a    506 c    800 g    528 t
```

```
ORIGIN


alignment_scores:
            Quality:  108.00              Length:    273
              Ratio:  0.831                Gaps:      6
 Percent Similarity:  47.619   Percent Identity: 17.216

alignment_block:
 US-09-068-804-5 x HUMM4PRO   ..

 Align seg 1/1  to: HUMM4PRO  from: 1  to: 2457

      10 MetGlyCysIleGlyLysValLeuGlyAlaLeuLeuThrIleValSerVa 26
         ||||||   |||:::   :::|||:::::::   |||:::        ::
    1113 ATGGGCAGGATAAATGAAATCCTAAGTAATGCACTGAAGAGAGGAGAGAT 1162

      26 lValAlaAlaValPheThrGlyGlyAlaSerLeuAlaLeuAlaAlaValG 43
         :::::|||           ||||||::::::   ::::::::::::::::
    1163 CATTGCAAAGCAGGGAGGAGGTGGAGGTGGAGGAAGCGTCCCTGGGATC. 1211

      43 lyLeuAlaValMetValAlaAspGluIleValLysAlaAlaThrGlyVal 59

    1211 ................................................... 1211

      60 SerPheIleGlnGlnAlaLeuAsnProIleMetGluHisValLeuLysPr 76
                    :::   ::::::|||   ::::::::::::      ::
    1212 ............GAGAGGATGGGTCCTGGCATTGACCGCCTCGGGGGTGC 1249

      76 oLeuMetGluLeuIleGlyLysAlaIleThrLysAlaLeuGluGlyLeuG 93
         :   ||||||   ::::||   ::::::      :::::::::   :::|
    1250 CGGCATGGAGCGCATGGGCGCGGGCCTGGGCCACGGCATGGATCGCGTGG 1299

      93 lyValAspArgLysArgGlnArgTrpProAlaAlaLeuLeuValArgLeu 109
         ||   :::      |||
    1300 GCTCCGAGATCGAGCGC.............................. 1316

     110 SerProLeuCysHisGlyAspAlaValIleValValValAlaValValGl 126

    1316 ................................................... 1316

     126 yLysGlyAlaAlaAlaLysLeuGlyAsnAlaLeuSerLysMet....... 140
                        :::|||   :::::::::::::|||
    1317 ...................ATGGGCCTGGTCATGGACCGCATGGGCTCCG 1347

     141 ........MetGlyGluThrIleLysLysLeuValProAsnValLeuLys 154
                 ||||||      |||   ::::::   |||      |||
    1348 TGGAGCGCATGGGCTCCGGCATTGAGCGCATGGGCCCGCTGGGCCTCGAC 1397

     155 GlnLeuAlaGlnAsnGlySerLysLeuPheThrGlnGlyMetGlnArgIl 171
         :::::::|||   :::      ::::::      |||   |||:::|||||
    1398 CACATGGCCTCCAGCATTGAGCGCATG...GGCCAGACCATGGAGCGCAT 1444

     171 eThrSerGlyLeuGlyAsnValGlySerLysMet......GlyLeuGlnT 186
         |   ||||||:::      :::|||:::   |||      ||||||:::
    1445 TGGCTCTGGCGTGGAGCGCATGGGTGCCGGCATGGGCTTCGGCCTTGAGC 1494

     186 hrAsnAlaLeuSerLysGluLeuValGlyAsnThrLeuAsnLysValAla 202
           |||   :::   :::   ||||||:::|||:::::::::::::::::
    1495 GCATGGCCGCTCCCATCGACCGTGTGGGCCAGACCATTGAGCGCATGGGC 1544

     203 LeuGlyMetGluValThrAsnThrAlaAlaGlnSerAlaGlyGlyValAl 219
         |||:::|||      ::::::|||   :::      |||
    1545 TCTGGCGTGGAGCGCATGGGCCCTGCCATCGAGCGCATGGGCCCGAGCAT 1594

     219 aGluGlyValPheIleLysAsnAlaSerGluAlaLeuAlaAspPheMetL 236
          |||   :::         :::   :::::::::|||:::
    1595 GGAGCGCATGGTGCCCGCAGGTATGGGAGCTGGCCTGGAG.......... 1634

     236 euAlaArgPheAlaMetAspGlnIleGlnGlnTrpLeuLysGlnSerVal 252
                       |||:::::::::   :::::::::      ::::::
    1635 ...........CGCATGGGCCCCGTGATGGATCGCATGGCCACCGGCCTG 1673

      253 GluIlePheGlyGluAsnGlnLysValThrAlaGluLeuGlnLysAlaMe 269
          |||       |||:::|||:::              |||::::::
     1674 GAGCGCATGGGCGCCAACAATCTGGAGCGGATGGGCCTGGAGCGCATGGG 1723

      269 tSerSerAlaValGlnGln 275
          ::::::::::::::::::
     1724 CGCCAACAGCCTCGAGCGC 1742

seq_name: gb_pat:A27284

seq_documentation_block:
LOCUS       A27284       1506 bp    DNA             PAT       17-OCT-1995
DEFINITION  H.sapiens TGR-CLH.
ACCESSION   A27284
NID         g1247482
VERSION     A27284.1  GI:1247482
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1506)
  AUTHORS   .
  TITLE     THYROIDAL RECEPTOR FOR N-ACETYLGLUCOSAMINE
  JOURNAL   Patent: WO 9219733-A 27 12-NOV-1992;
FEATURES             Location/Qualifiers
     source          1. .1506
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     gene            1. .1011
                     /gene="TGR-CLH"
     CDS             <1. .1011
                     /gene="TGR-CLH"
                     /codon_start=1
                     /protein_id="CAA01873.1"
                     /db_xref="PID:e205123"
                     /db_xref="PID:g1247483"
                     /db_xref="GI:1247483"
                     /translation="EFGASMGCIGPAGMGMEGIGFGINKMGGMEGPFGGGMENMGRFG
                     SGMNMGRINEILSNALKRGEIIAKQGGGGGGGSVPGIERMGPGIDRLGGAGMERMGAG
                     LGHGMDRVGSEIERMGLVMDRMGSVERMGSGIERMGPLGLDHMASSIERMGQTMERIG
                     SGVERMGAGMGFGLERMAAPIDRVGQTIERMGSGVERMGPAIERMGLSMERMVPAGMG
                     AGLERMGPVMDRMPPAWSAWAPTIWSGWAWSAWAPTASSAWAWSAWVPTASSAWAPPW
                     ARPWALALSAWAWPWVAVAVPALTVPSRWSVATSEEASQVPLVELEAMLLGWPGRPAR
                     YL"
BASE COUNT      328 a    324 c    520 g    334 t
ORIGIN


alignment_scores:
            Quality:  105.00              Length:    242
              Ratio:  0.882                Gaps:     11
 Percent Similarity:  49.174   Percent Identity: 20.661

alignment_block:
 US-09-068-804-5 x A27284   ..

 Align seg 1/1  to: A27284  from: 1  to: 1506

      10 MetGlyCysIleGlyLysValLeuGlyAlaLeuLeuThrIleValSerVa 26
         ||||||||||||
      16 ATGGGTTGCATA..................................... 27

      26 lValAlaAlaValPheThrGlyGlyAlaSerLeuAlaLeuAlaAlaValG 43
                            |||   |||:::::::::::::::::::::::|
      28 ...................GGTCCCGCAGGAATGGGAATGGAAGGCATAG 58

      43 lyLeuAlaVal...............MetValAla.............. 49
         ||:::::::::               |||   :::
      59 GATTTGGAATAAATAAAATGGGAGGAATGGAGGGGCCCTTTGGTGGTGGT 108

      50 ...AspGluIleValLysAlaAlaThrGlyValSerPheIleGlnGlnAl 65
            :::::::::   :::   ::::::|||::::::      :::
```

```
    109 ATGGAAAACATGGGTCGATTTGGATCTGGGATGAACATGGGCAGG..... 153

     65 aLeuAsnProIleMetGluHisValLeuLys..................... 75
        :::|||   |||:::   ::::::||||||
    154 .ATAAATGAAATCCTAAGTAATGCACTGAAGAGAGGAGAGATCATTGCAA 202

     76 ..............................ProLeuMetGluLeuIle 81
                                       |||   :::|||   :::
    203 AGCAGGGAGGAGGTGGAGGTGGAGGAAGCGTCCCTGGGATCGAGAGGATG 252

     82 GlyLysAlaIleThrLysAlaLeuGluGlyLeuGlyValAspArgLysAr 98
        |||   :::|||   :::   |||   |||   |||::::::|||:::
    253 GGTCCTGGCATTGACCGC...CTCGGGGGTGCCGGCATGGAGCGCATG.. 297

     98 gGlnArgTrpProAlaAlaLeuLeuValArgLeuSerProLeuCysHisG 115
                                                         |
    298 ..............................................G 298

    115 lyAspAlaValIleValValValAlaValValGlyLysGlyAlaAlaAla 131
        ||:::::::::          ::::::   ||||||   ::::::   :::
    299 GCGCGGGCCTGGGCCACGGCATGGATCGCGTGGGC...TCCGAGATCGAG 345

    132 LysLeuGlyAsnAlaLeuSerLysMet..............MetGlyGl 143
        :::::::|||   :::::::::::::|||              ||||||
    346 CGCATGGGCCTGGTCATGGACCGCATGGGCTCCGTGGAGCGCATGGGCTC 395

    143 uThrIleLysLysLeuValProAsnValLeuLysGlnLeuAlaGlnAsnG 160
           |||   ::::::   |||      |||   ::::::|||   :::
    396 CGGCATTGAGCGCATGGGCCCGCTGGGCCTCGACCACATGGCCTCCAGCA 445

    160 lySerLysLeuPheThrGlnGlyMetGlnArgIleThrSerGlyLeuGly 176
           ::::::      |||   |||:::|||||   ||||||:::
    446 TTGAGCGCATG...GGCCAGACCATGGAGCGCATTGGCTCTGGCGTGGAG 492

    177 AsnValGlySerLysMet......GlyLeuGlnThrAsnAlaLeuSerLy 191
           :::|||:::   |||      |||||::      |||   :::
    493 CGCATGGGTGCCGGCATGGGCTTCGGCCTTGAGCGCATGGCCGCTCCCAT 542

    191 sGluLeuValGlyAsnThrLeuAsnLysValAlaLeuGlyMetGluValT 208
         :::   ||||||:::|||:::::::::::::::   |||:::|||
    543 CGACCGTGTGGGCCAGACCATTGAGCGCATGGGCTCTGGCGTGGAGCGCA 592

    208 hrAsnThrAlaAlaGlnSerAlaGly 216
        ::::::|||   :::      |||
    593 TGGGCCCTGCCATCGAGCGCATGGGC 618
```

```
seq_name: gb_pr2:HSTGR1

seq_documentation_block:
LOCUS       HSTGR1      2109 bp    mRNA            PRI       19-FEB-1996
DEFINITION  H.sapiens hTGR 1 mRNA.
ACCESSION   X72018
NID         g1200088
VERSION     X72018.1  GI:1200088
KEYWORDS    monomer; N-acetylglucosamine; receptor; transmembrane protein.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2109)
  AUTHORS   Blanck,O., Perrin,C., Mziaut,H., Darbon,H., Mattei,M.G. and
            Miquelis,R.
  TITLE     Molecular cloning, cDNA analysis, and localization of a monomer of
            the N-acetylglucosamine-specific receptor of the thyroid, NAGR1, to
            chromosome 19p13.3-13.2
  JOURNAL   Genomics 21 (1), 18-26 (1994)
  MEDLINE   94375011
  REMARK    Erratum:[[published erratum appears in Genomics 1995 Jun
            10;27(3):561]]
REFERENCE   2  (bases 1 to 2106)
  AUTHORS   Miquelis,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-1993) R. Miquelis, CNRS, Laboratoire de
            Biochimie. URA CNRS 1455, Faculte de Medicine Nord, Bd. Pierre
            Dramard, 13915 Marseille Cedex 15, FRANCE
REFERENCE   3  (bases 1 to 2106)
  AUTHORS   Blanck,O., Perrin,C., Mziaut,H., Darbon,H., Mattei,M.G. and
            Miquelis,R.
  TITLE     Molecular cloning, cDNA analysis, and localization of a monomer of
            the N-acetylglucosamine-specific receptor of the thyroid, NAGR1, to
            chromosome 19p13.3-13.2
  JOURNAL   Genomics 27 (3), 561 (1995)
  MEDLINE   96047351
COMMENT     On Feb 23, 1996 this sequence version replaced gi:417613.
FEATURES             Location/Qualifiers
     source          1. .2109
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="19p 13.2-13.3"
                     /dev_stage="adult"
                     /tissue_type="thyroid"
     gene            127. .1839
                     /gene="ORF"
     CDS             127. .1839
                     /gene="ORF"
                     /note="almost identical to nRNP M protein, acc.L03532"
                     /codon_start=1
                     /protein_id="CAA50897.1"
                     /db_xref="PID:g1200089"
                     /db_xref="GI:1200089"
                     /db_xref="SPTREMBL:Q15584"
                     /translation="MATTGGMGMGPGGPGMITIPPSILNNPNIPNEIIHALQAGRLGS
                     TVFVANLDYKVGWKKLKEVFSMAGVVVRADILEDKDGKSRGIGTVTFEQSIEAVQAIS
                     MFNGQLLFDRPMHVKMDERALPKGDFFPPERPQQLPHGLGGIGMGLGPGGQPIDANHL
                     NKGIGMGNIGPAGMGMEGIGFGINKMGGMEGPFGGGMENMGRFGSGMNMGRINEILSN
                     ALKRGEIIAKQGGGGGGGSVPGIERMGPGIDRLGGAGMERMGAGLGHGMDRVGSEIER
                     MGLVMDRMGSVERMGSGIERMGPLGLDHMASSIERMGQTMERIGSGVERMGAGMGFGL
                     ERMAAPIDRVGQTIERMGSGVERMGPAIERMGLSMERMVPAGMGAGLERMGPVMDRMA
                     TGLERMGANNLERMGLERMGANSLERMGLERMGANSLERMGPAMGPALGAGIERMGLA
                     MGGGGGASFDRAIEMERGNFGGSFAGSFGGAGGHAPGVARKACQIFVRNLPFDFTWKM
                     LKDKFNECGHVLYADIKMENGKSKGCGVVKFESPEVAERACRMMNGMKLSGREIDVRI
                     DRNA"
     polyA_signal    2078. .2083
BASE COUNT      517 a    443 c    678 g    471 t
ORIGIN


alignment_scores:
            Quality:  106.00              Length:    297
              Ratio:  0.752                 Gaps:      9
 Percent Similarity:  47.475   Percent Identity: 16.498

alignment_block:
 US-09-068-804-5 x HSTGR1   ..

 Align seg 1/1  to: HSTGR1  from: 1  to: 2109

     10 MetGlyCysIleGlyLysValLeuGlyAlaLeuLeuThrIleValSerVa 26
        ||||||   |||:::   :::|||:::::   |||:::          ::
    751 ATGGGCAGGATAAATGAAATCCTAAGTAATGCACTGAAGAGAGGAGAGAT 800

     26 lValAlaAlaValPheThrGlyGlyAlaSerLeuAlaLeuAlaAlaValG 43
        :::|||            ||||||::::::   ::::::::::::::
    801 CATTGCAAAGCAGGGAGGAGGTGGAGGTGGAGGAAGCGTCCCTGGGATC. 849

     43 lyLeuAlaValMetValAlaAspGluIleValLysAlaAlaThrGlyVal 59

    849 ..................................................... 849

     60 SerPheIleGlnGlnAlaLeuAsnProIleMetGluHisValLeuLysPr 76
                ::   ::::::|||   ::::::::::::      ::
    850 ...........GAGAGGATGGGTCCTGGCATTGACCGCCTCGGGGGTGC 887

     76 oLeuMetGluLeuIleGlyLysAlaIleThrLysAlaLeuGluGlyLeuG 93
        :   ||||||   :::|||   ::::::      :::::::::   :::|
    888 CGGCATGGAGCGCATGGGCGCGGGCCTGGGCCACGGCATGGATCGCGTGG 937
```

```
     93 lyValAspArgLysArgGlnArgTrpProAlaAlaLeuLeuValArgLeu 109
        ||   :::     |||
    938 GCTCCGAGATCGAGCGC.............................. 954

    110 SerProLeuCysHisGlyAspAlaValIleValValValAlaValValGl 126

    954 ..................................................... 954

    126 yLysGlyAlaAlaAlaLysLeuGlyAsnAlaLeuSerLysMet....... 140
                          :::|||   :::::::::::::||
    955 ..................ATGGGCCTGGTCATGGACCGCATGGGCTCCG 985

    141 ........MetGlyGluThrIleLysLysLeuValProAsnValLeuLys 154
                |||||      |||   ::::::    |||       |||
    986 TGGAGCGCATGGGCTCCGGCATTGAGCGCATGGGCCCGCTGGGCCTCGAC 1035

    155 GlnLeuAlaGlnAsnGlySerLysLeuPheThrGlnGlyMetGlnArgIl 171
        ::::::|||   :::       ::::::      |||   |||:::||||
   1036 CACATGGCCTCCAGCATTGAGCGCATG...GGCCAGACCATGGAGCGCAT 1082

    171 eThrSerGlyLeuGlyAsnValGlySerLysMet......GlyLeuGlnT 186
        |   ||||||:::       :::|||:::   |||       ||||||:::
   1083 TGGCTCTGGCGTGGAGCGCATGGGTGCCGGCATGGGCTTCGGCCTTGAGC 1132

    186 hrAsnAlaLeuSerLysGluLeuValGlyAsnThrLeuAsnLysValAla 202
          |||   :::   :::   ||||||:::|||:::::::::::::::::
   1133 GCATGGCCGCTCCCATCGACCGTGTGGGCCAGACCATTGAGCGCATGGGC 1182

    203 LeuGlyMetGluValThrAsnThrAlaAlaGlnSerAlaGly........ 216
        |||:::|||       :::::::|||   :::       |||
   1183 TCTGGCGTGGAGCGCATGGGCCCTGCCATCGAGCGCATGGGCCTGAGCAT 1232

    217 ..................GlyValAlaGluGlyVal.............. 222
                           |||:::::::::|||:::
   1233 GGAGCGCATGGTGCCCGCAGGTATGGGAGCTGGCCTGGAGCGCATGGGCC 1282

    223 ........................PheIleLysAsnAlaSerGluAla 230
                               :::   :::   ::::::::::::
   1283 CCGTGATGGATCGCATGGCCACCGGCCTGGAGCGCATGGGCGCCAACAAT 1332

    231 LeuAlaAspPheMetLeuAlaArgPheAlaMetAspGlnIleGlnGlnTr 247
        |||:::         |||:::|||   :::   :::   ::::::
   1333 CTGGAGCGGATGGGCCTGGAGCGCATGGGCGCCAACAGCCTCGAG..... 1377

    247 pLeuLysGlnSerValGluIlePheGlyGluAsn......GlnLysValT 262
          :::   ::::::|||      |||:::|||      :::::::::
   1378 ....CGCATGGGCCTGGAGCGCATGGGTGCCAACAGCCTCGAGCGCATGG 1423

    262 hrAlaGluLeuGlnLysAlaMetSerSerAlaValGlnGln 275
        :::::::::      |||:::::::::::::::::::::
   1424 GCCCCGCCATGGGCCCGGCCCTGGGCGCTGGCATTGAGCGC 1464

seq_name: gb_pat:E12520

seq_documentation_block:
LOCUS       E12520       1464 bp    DNA             PAT       24-JUN-1998
DEFINITION  DNA encoding an antigen peptide of Chlamydia pneumoniae.
ACCESSION   E12520
NID         g3251353
VERSION     E12520.1  GI:3251353
KEYWORDS    JP 1997009974-A/1.
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1464)
  AUTHORS   Izutsu,H. and Matsumoto,A. .
  TITLE     ANTIGEN POLYPEPTIDE INVOLVED WITH CHLAMYDIA PNEUMONEAE, DNA CODING
            FOR THE SAME, RECOMBINANT VECTOR CONTAINING THE DNA, TRANSFORMANT
            CONTAINING THE VECTOR, AND PRODUCTION OF ANTI-CHLAMYDIA PNEUMONEAE
  JOURNAL   Patent: JP 1997009974-A 1 14-JAN-1997;
            HITACHI CHEM CO LTD
COMMENT     OS   Chlamydia pneumoniae
            PN   JP 1997009974-A/1
            PD   14-JAN-1997
            PF   20-SEP-1995 JP 1995242095
            PR   20-SEP-1994 JP 94P   224711, 28-APR-1995 JP 95P   106010 PI
            IZUTSU HIROSHI, MATSUMOTO AKIRA
            PC   C12N15/09,C07H21/04,C07K14/295,C07K16/12,C12N1/21,C12N15/02,
            PC   C12P21/08,
            PC   G01N33/53,G01N33/571//A61K39/118,A61K39/118,A61K39/118, PC
            (C12N1/21,C12R1:19),
            PC   (C12P21/08,C12R1:91);
            CC   strandedness: Double;
            CC   topology: Linear;
            FH   Key             Location/Qualifiers
            FH
            FT   source          1. .1464
            FT                   /organism='Chlamydia pneumoniae' FT
                   /strain='YK41'
            FT   mat_peptide     1. .1464
            FT                   /product='antigen peptide'.
FEATURES                Location/Qualifiers
     source             1. .1464
                        /organism="unidentified"
                        /db_xref="taxon:32644"
BASE COUNT      470 a    331 c    348 g    315 t
ORIGIN


alignment_scores:
            Quality:  101.50            Length:    227
              Ratio:  0.839               Gaps:      6
 Percent Similarity:  53.304   Percent Identity: 23.789

alignment_block:
 US-09-068-804-5 x E12520   ..

 Align seg 1/1  to: E12520  from: 1  to: 1464

      1 ArgLysAlaGluGluThrAsnArgIleMetGlyCysIleGlyLysValLe 17
        :::|||:::   :::   :::   :::|||:::   ::::::::::|||::
    721 CAGAAGTCTAAAGATCTCGAAGGAACAATGGATACTGTCAATACTGTGAT 770

     17 uGlyAlaLeu.........LeuThrIleValSerValValAlaAlaValP 31
        :  |||:::          :::|||:::::::|||:::|||||||||:::|
    771 GATCGCGGTTTCTGTTGCCATTACAGTTATTTCTATTGTTGCTGCTATTT 820

     31 heThrGlyGlyAlaSerLeu............AlaLeuAlaAlaValGly 43
        |||||   ||||||:::|||            |||   ||||||||||||
    821 TTACATGCGGAGCTGGACTCGCTGGACTCGCTGCGGGAGCTGCTGTAGGT 870

     44 LeuAlaValMetValAlaAspGluIleValLysAlaAlaThrGlyValSe 60
           |||:::      :::::::::   :::   |||||||||   |||::
    871 GCAGCGGCAGCTGGAGGTGCAGCAGGAGCTGCTGCCGCAACCACGGTAGC 920

     60 rPheIleGlnGlnAlaLeuAsnProIleMetGluHisValLeuLysProL 77
        :   :::|||   :::::::::::::::::::::   |||   :::::::
    921 A...ACACAAATTACAGTTCAAGCTGTTGTCCAAGCGGTGAAACAAGCTG 967

     77 euMetGluLeuIleGlyLysAlaIleThrLysAlaLeuGluGlyLeuGly 93
        :::::      :::   :::||||||||||   |||:::
    968 TTATCACAGCTGTCAGACAAGCGATCACCGCGGCTATAAAA......... 1008

     94 ValAspArgLysArgGlnArgTrpProAlaAlaLeuLeuValArgLeuSe 110

   1008 ..................................................... 1008

    110 rProLeuCysHisGlyAspAlaValIleValValValAlaValValGlyL 127

   1008 ..................................................... 1008

    127 ysGlyAlaAlaAlaLysLeuGlyAsnAlaLeuSerLysMetMetGlyGlu 143
              ||||||:::|||   |||                   :::   :::
   1009 .....GCGGCTGTCAAATCTGGA.................ATAAAAGCA 1035
```

```
     144 ThrIleLysLysLeuValProAsnValLeuLysGlnLeuAlaGlnAsnGl 160
         ||||||:::|||||   ::::::   |||   ::::::::::::
    1036 TTTATCAAAACTTTAGTCAAAGCGATTGCCAAAGCCATTTCTAAAGGAAT 1085

     160 ySerLysLeuPheThrGlnGlyMetGlnArgIleThrSerGlyLeuGlyA 177
         ||||||:::|||:::::|||   |||   |||:::   ::::::   :
    1086 CTCTAAGGTTTTCGCTAAGGGAACTCAAATGATTGCGAAGAACTTCCCCA 1135

     177 snValGlySerLysMetGlyLeuGlnThrAsnAlaLeuSerLysGluLeu 193
         ::                                |||||||||   :::
    1136 AG..............................CTCTCGAAAGTCATC 1152

     194 ValGlyAsnThrLeuAsnLysValAlaLeuGlyMetGluValThrAsnTh 210
         :::   |||   :::   |||:::::::|||:::   |||:::   ::
    1153 TCGTCTCTTACCAGTAAATGGGTCACGGTTGGGGTTGGGGTTGTAGTTGC 1202

     210 rAlaAlaGlnSerAlaGlyGlyValAlaGlu 220
         :|||:::      :::   |||:::   :::
    1203 GGCGCCTGCTCTCGGTAAAGGGATTATGCAA 1233

seq_name: gb_pat:E12535

seq_documentation_block:
LOCUS       E12535       1464 bp    DNA             PAT       24-JUN-1998
DEFINITION  Chlamydia pneumoniae 53KD-antigen gene.
ACCESSION   E12535
NID         g3251368
VERSION     E12535.1  GI:3251368
KEYWORDS    JP 1997009999-A/1.
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1464)
  AUTHORS   Izutsu,H. and Matsumoto,A. .
  TITLE     PROBE AND PRIMER FOR MEASURING CHLAMYDIA PNEUMONIAE GENE,
            MEASUREMENT OF CHLAMYDIA PNEUMONIAE GENE USING THE SAME PROBE OR
            PRIMER AND REAGENT FOR MEASURING CHLAMYDIA PENUMONIAE GENE
  JOURNAL   Patent: JP 1997009999-A 1 14-JAN-1997;
            HITACHI CHEM CO LTD
COMMENT     OS   Chlamydia pneumoniae
            PN   JP 1997009999-A/1
            PD   14-JAN-1997
            PF   15-MAR-1996 JP 1996058608
            PR   28-APR-1995 JP 95P   106008
            PI   IZUTSU HIROSHI, MATSUMOTO AKIRA
            PC   C12Q1/68,C07H21/04,C12N1/21,C12N15/02,C12N15/09,C12P21/08, PC
             G01N33/566,
            PC   G01N33/569,G01N33/571//A61K49/00,C07K14/295,C12Q1/04,
            (C12N1/21, PC   C12R1:19);
            CC   strandedness: Double;
            CC   topology: Linear;
            FH   Key             Location/Qualifiers
            FH
            FT   source          1. .1464
            FT                   /organism='Chlamydia pneumoniae' FT
                   /strain='YK-41'
            FT   mat_peptide     1. .1464
            FT                   /product='53KD-antigen peptide'.
FEATURES             Location/Qualifiers
     source          1. .1464
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      470 a    331 c    348 g    315 t
ORIGIN


alignment_scores:
            Quality:  101.50            Length:    227
              Ratio:  0.839               Gaps:      6
 Percent Similarity:  53.304   Percent Identity: 23.789

alignment_block:
 US-09-068-804-5 x E12535   ..
```

```
Align seg 1/1  to: E12535  from: 1  to: 1464

       1 ArgLysAlaGluGluThrAsnArgIleMetGlyCysIleGlyLysValLe 17
         :::|||:::   :::   :::   :::|||:::   :::::::::|||::
     721 CAGAAGTCTAAAGATCTCGAAGGAACAATGGATACTGTCAATACTGTGAT 770

      17 uGlyAlaLeu.........LeuThrIleValSerValValAlaAlaValP 31
         :   |||:::          :::|||:::::::|||:::|||||||||:::|
     771 GATCGCGGTTTCTGTTGCCATTACAGTTATTTCTATTGTTGCTGCTATTT 820

      31 heThrGlyGlyAlaSerLeu.............AlaLeuAlaAlaValGly 43
         |||||   ||||||:::|||             |||   ||||||||||||
     821 TTACATGCGGAGCTGGACTCGCTGGACTCGCTGCGGGAGCTGCTGTAGGT 870

      44 LeuAlaValMetValAlaAspGluIleValLysAlaAlaThrGlyValSe 60
         |||:::      :::::::::   :::   |||||||||   |||::
     871 GCAGCGGCAGCTGGAGGTGCAGCAGGAGCTGCTGCCGCAACCACGGTAGC 920

      60 rPheIleGlnGlnAlaLeuAsnProIleMetGluHisValLeuLysProL 77
         :   :::|||   :::::::::::::::::::::   |||   :::::::
     921 A...ACACAAATTACAGTTCAAGCTGTTGTCCAAGCGGTGAAACAAGCTG 967

      77 euMetGluLeuIleGlyLysAlaIleThrLysAlaLeuGluGlyLeuGly 93
         :::::      :::  :::||||||||||   |||:::
     968 TTATCACAGCTGTCAGACAAGCGATCACCGCGGCTATAAAA..... 1008

      94 ValAspArgLysArgGlnArgTrpProAlaAlaLeuLeuValArgLeuSe 110

    1008 ..................................................... 1008

     110 rProLeuCysHisGlyAspAlaValIleValValValAlaValValGlyL 127

    1008 ..................................................... 1008

     127 ysGlyAlaAlaAlaLysLeuGlyAsnAlaLeuSerLysMetMetGlyGlu 143
          ||||||:::|||   |||                  :::   :::
    1009 .....GCGGCTGTCAAATCTGGA.................ATAAAAGCA 1035

     144 ThrIleLysLysLeuValProAsnValLeuLysGlnLeuAlaGlnAsnGl 160
         ||||||:::||||||   ::::::   |||   ::::::::::::
    1036 TTTATCAAAACTTTAGTCAAAGCGATTGCCAAAGCCATTTCTAAAGGAAT 1085

     160 ySerLysLeuPheThrGlnGlyMetGlnArgIleThrSerGlyLeuGlyA 177
         ||||||:::|||:::::|||   |||   |||:::   ::::::   :
    1086 CTCTAAGGTTTTCGCTAAGGGAACTCAAATGATTGCGAAGAACTTCCCCA 1135

     177 snValGlySerLysMetGlyLeuGlnThrAsnAlaLeuSerLysGluLeu 193
         ::                                |||||||||   :::
    1136 AG..............................CTCTCGAAAGTCATC 1152

     194 ValGlyAsnThrLeuAsnLysValAlaLeuGlyMetGluValThrAsnTh 210
         :::   |||   :::   |||:::::::|||:::   |||:::   ::
    1153 TCGTCTCTTACCAGTAAATGGGTCACGGTTGGGGTTGGGGTTGTAGTTGC 1202

     210 rAlaAlaGlnSerAlaGlyGlyValAlaGlu 220
         :|||:::      :::   |||:::   :::
    1203 GGCGCCTGCTCTCGGTAAAGGGATTATGCAA 1233

seq_name: gb_pat:E12541

seq_documentation_block:
LOCUS       E12541       1464 bp    DNA             PAT       24-JUN-1998
DEFINITION  DNA encoding an antigen peptide Chlamydia pneumoniae.
ACCESSION   E12541
NID         g3251374
VERSION     E12541.1  GI:3251374
KEYWORDS    JP 1997015244-A/1.
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1464)
  AUTHORS   Izutsu,H. and Matsumoto,A. .
```

```
  TITLE     MEASURING METHOD FOR ANTICHLAMYDIA-PNEUMONIA ANTIBODY AND REAGENT
            THEREFOR AS WELL AS DIAGNOSTIC MEDICINE FOR CHLAMYDIA PNEUMONIA
            INFECTION
  JOURNAL   Patent: JP 1997015244-A 1 17-JAN-1997;
            HITACHI CHEM CO LTD
COMMENT     OS   Chlamydia pneumoniae
            PN   JP 1997015244-A/1
            PD   17-JAN-1997
            PF   14-MAR-1996 JP 1996057410
            PR   28-APR-1995 JP 95P   106011
            PI   IZUTSU HIROSHI, MATSUMOTO AKIRA
            PC   G01N33/571,C07K14/295,C07K16/12,G01N33/53,
            PC   G01N33/569//A61K49/00,C12N15/09;
            CC   strandedness: Double;
            CC   topology: Linear;
            FH   Key             Location/Qualifiers
            FH
            FT   source          1. .1464
            FT                   /organism='Chlamydia pneumoniae' FT
              /strain='YK41'
            FT   mat_peptide     1. .1464
            FT                   /product='antigen peptide'.
FEATURES            Location/Qualifiers
     source         1. .1464
                    /organism="unidentified"
                    /db_xref="taxon:32644"
BASE COUNT      470 a    331 c    348 g    315 t
ORIGIN


alignment_scores:
            Quality:  101.50              Length:    227
              Ratio:  0.839                 Gaps:      6
 Percent Similarity:  53.304   Percent Identity: 23.789

alignment_block:
 US-09-068-804-5 x E12541   ..

 Align seg 1/1  to: E12541  from: 1  to: 1464

       1 ArgLysAlaGluGluThrAsnArgIleMetGlyCysIleGlyLysValLe 17
         :::|||:::   :::   :::   :::|||:::   :::::::::|||:: 
     721 CAGAAGTCTAAAGATCTCGAAGGAACAATGGATACTGTCAATACTGTGAT 770

      17 uGlyAlaLeu.........LeuThrIleValSerValValAlaAlaValP 31
         :   |||:::         :::|||::::::|||:::|||||||||:::| 
     771 GATCGCGGTTTCTGTTGCCATTACAGTTATTTCTATTGTTGCTGCTATTT 820

      31 heThrGlyGlyAlaSerLeu.............AlaLeuAlaAlaValGly 43
         |||||   ||||||:::|||             |||   ||||||||||| 
     821 TTACATGCGGAGCTGGACTCGCTGGACTCGCTGCGGGAGCTGCTGTAGGT 870

      44 LeuAlaValMetValAlaAspGluIleValLysAlaAlaThrGlyValSe 60
         |||:::      :::::::::   :::   |||||||||   |||:: 
     871 GCAGCGGCAGCTGGAGGTGCAGCAGGAGCTGCTGCCGCAACCACGGTAGC 920

      60 rPheIleGlnGlnAlaLeuAsnProIleMetGluHisValLeuLysProL 77
         :   :::|||   :::::::::::::::::::::   |||   ::::::: 
     921 A...ACACAAATTACAGTTCAAGCTGTTGTCCAAGCGGTGAAACAAGCTG 967

      77 euMetGluLeuIleGlyLysAlaIleThrLysAlaLeuGluGlyLeuGly 93
         :::::      :::   :::|||||||||   |||:::
     968 TTATCACAGCTGTCAGACAAGCGATCACCGCGGCTATAAAA..... 1008

      94 ValAspArgLysArgGlnArgTrpProAlaAlaLeuLeuValArgLeuSe 110

    1008 ..................................................... 1008

     110 rProLeuCysHisGlyAspAlaValIleValValValAlaValValGlyL 127

    1008 ..................................................... 1008

     127 ysGlyAlaAlaAlaLysLeuGlyAsnAlaLeuSerLysMetMetGlyGlu 143

         ||||||:::|||   |||                  :::   :::
    1009 .....GCGGCTGTCAAATCTGGA.................ATAAAAGCA 1035

     144 ThrIleLysLysLeuValProAsnValLeuLysGlnLeuAlaGlnAsnGl 160
         |||||||:::||||||   ::::::   |||   ::::::::::::
    1036 TTTATCAAAACTTTAGTCAAAGCGATTGCCAAAGCCATTTCTAAAGGAAT 1085

     160 ySerLysLeuPheThrGlnGlyMetGlnArgIleThrSerGlyLeuGlyA 177
         ||||||:::|||:::::::|||   |||   |||:::   ::::::   :
    1086 CTCTAAGGTTTTCGCTAAGGGAACTCAAATGATTGCGAAGAACTTCCCCA 1135

     177 snValGlySerLysMetGlyLeuGlnThrAsnAlaLeuSerLysGluLeu 193
         ::                                  |||||||||   :::
    1136 AG..............................CTCTCGAAAGTCATC 1152

     194 ValGlyAsnThrLeuAsnLysValAlaLeuGlyMetGluValThrAsnTh 210
         :::   |||   :::   |||::::::|||:::   |||:::   ::
    1153 TCGTCTCTTACCAGTAAATGGGTCACGGTTGGGGTTGGGGTTGTAGTTGC 1202

     210 rAlaAlaGlnSerAlaGlyGlyValAlaGlu 220
         :|||:::      :::   |||:::   :::
    1203 GGCGCCTGCTCTCGGTAAAGGGATTATGCAA 1233
```

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 1999, 14:49:17 ; Search time 1495.71 Seconds
(without alignments)
1147.351 Million cell updates/sec

Title: US-09-068-804-1
Perfect score: 870
Sequence: 1 CGCAAAGCCGAGGAAACGAA.....TGCGCCAGAGTCGCGCATAA 870

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 986266752 residues

Database : EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		55.2	6.3	408	34	AA484400	AA484400 nf10e04.s
2		46	5.3	527	43	AI238960	AI238960 GH15115.5
3		42.8	4.9	686	23	D22789	D22789 RICC1332A R
4	c	41.4	4.8	554	51	AI728199	AI728199 BNLGHi101
5	c	41.2	4.7	628	51	AI731721	AI731721 BNLGHi105
6	c	40.6	4.7	590	44	AI292820	AI292820 GH15821.5
7	c	40.2	4.6	487	38	AA749517	AA749517 SWAMCAC06
8	c	39	4.5	650	33	AA390540	AA390540 LD08780.5
9	c	39	4.5	616	44	AI292805	AI292805 GH15806.5
10	c	39	4.5	468	44	AI292818	AI292818 GH15819.5
11	c	38.6	4.4	531	38	AA441313	AA441313 LD16203.5
12		38	4.4	564	24	H76570	H76570 18275 Lambd
13	c	37.8	4.3	619	41	AI067558	AI067558 EST209240
14		37.8	4.3	420	43	AU033900	AU033900 AU033900
15	c	37.8	4.3	676	45	AI389106	AI389106 GH20192.5
16		37.6	4.3	237	23	R30281	R30281 12886 Lambd
17		37.4	4.3	558	34	AA483903	AA483903 ne92a10.s
18	c	37.4	4.3	580	37	AA699167	AA699167 HL07844.5
19		37.4	4.3	537	43	AI237995	AI237995 GH13939.5
20		37.4	4.3	557	46	AI404106	AI404106 GH23705.5
21		37.2	4.3	756	35	C25541	C25541 C25541 Dict
22	c	37.2	4.3	492	51	AI737278	AI737278 606038D02
23		37	4.3	314	28	AA083950	AA083950 zn16e01.r
24		36.8	4.2	609	51	AU070274	AU070274 AU070274
25		36.4	4.2	474	40	AA979220	AA979220 LD33551.5
26		36.4	4.2	736	44	AU039498	AU039498 AU039498
27	c	36.2	4.2	250	28	AA061257	AA061257 ml39c12.r
28		36	4.1	435	25	N79538	N79538 zb08h10.s1
29		36	4.1	382	28	AA119823	AA119823 mn14h06.r
30	c	36	4.1	601	47	AI484260	AI484260 EST248487
31	c	35.8	4.1	413	20	T32231	T32231 EST45490 Hu
32		35.8	4.1	242	37	AA727427	AA727427 vu96a04.r
33		35.8	4.1	504	40	AA913867	AA913867 om21g09.s
34		35.8	4.1	749	40	AA942336	AA942336 LD26481.5
35		35.8	4.1	509	42	AI135323	AI135323 GH13009.5
36		35.8	4.1	484	44	AI299758	AI299758 qn03g05.x
37		35.8	4.1	467	45	AI372595	AI372595 EST175490
38	c	35.8	4.1	501	50	AI668376	*AI668376 605031E06*
39		35.6	4.1	697	44	AI259916	AI259916 LP03440.5
40		35.6	4.1	676	44	AI260133	AI260133 LP03825.5
41		35.6	4.1	646	44	AI296199	AI296199 LP10035.5
42		35.6	4.1	635	47	AI514057	AI514057 GH27111.5
43		35.6	4.1	489	47	AI533308	AI533308 SD05070.5
44		35.4	4.1	373	41	AI062693	AI062693 GH01983.5
45		35.4	4.1	497	43	AI238309	AI238309 GH14332.5

ALIGNMENTS

RESULT 1
AA484400
LOCUS AA484400 408 bp mRNA EST 18-AUG-1997
DEFINITION nf10e04.s1 NCI_CGAP_Li1 Homo sapiens cDNA clone IMAGE:913374 similar to contains element MSR1 MSR1 repetitive element ;, mRNA sequence.
ACCESSION AA484400
NID g2213213

```
VERSION     AA484400.1  GI:2213213
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 408)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On May 9, 1995 this sequence version replaced gi:803122.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: David E. Kleiner, M.D., Ph.D., Rodrigo F.
            Chuaqui, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 638   Std Error: 0.00
            Seq primer: -41ml3 fwd. ET from Amersham
            High quality sequence stop: 362.
FEATURES              Location/Qualifiers
     source           1. .408
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:913374"
                      /clone_lib="NCI_CGAP_Li1"
                      /tissue_type="liver"
                      /lab_host="DH10B"
                      /note="Vector: pAMP10; mRNA made from normal liver
                      hepatocytes, cDNA made by oligo-dT priming.
                      Non-directionally cloned. Size-selected on agarose gel,
                      average insert  size 600 bp. Reference:  Krizman et al.
                      (1996) Cancer Research 56:5380-5383."
BASE COUNT       64 a     17 c    208 g    119 t
ORIGIN
```

```
  Query Match             6.3%;  Score 55.2;  DB 34;  Length 408;
  Best Local Similarity   47.8%;  Pred. No. 6.6e-05;
  Matches  193;  Conservative    0;  Mismatches  208;  Indels    3;  Gaps    1;

Qy      46 GTCCTCGGCGCGCTGCTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTGGGGCG 105
           | | | |  | |  | |    || || |  |  ||||  |  ||| |    |  | | |
Db       5 GACATGGTGGTGGAGGTGGTGATGGTGATGGCGGTGGTGGTGGTTATGGTGGTGGAGATG 64

Qy     106 AGTCTGGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCG 165
           |||  | |||  |||| ||   | | ||| ||||||||  |  | | | ||||  | |
Db      65 GTGGTGGTGATGGTGGCGGAGGTGGTGGTGGTGATGGTGGCAGTGGCAGTGGTGATAGTG 124

Qy     166 GCGACGGGAGTGTCGTTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAG 225
             |  ||       | | |||  |  || | |     | | |||| |   | | ||  |
Db     125 ATGGAGGTGACAATGGTCATTGTGGTGGTGGTTGTTGTGGTAATGGTGGTGGAGGTGGTG 184

Qy     226 CCGTTAATGGAGCTGATTGGCAAGGCGATTACCAAAGCGCTGGAAGGATTAGGCGTCGAT 285
             | | ||||   || | |    || |||    |  |   |||  | | | |   | |
Db     185 ATGGTGATGGTAGTGGTGGCAGTGGTGATGGTGATGGTAGTGGTGGCAGTGGTGATAGTG 244

Qy     286 AAGAAAACGGCAGAGATGGCCGGCAGCATTGTTGGTGCGATTGTCGCCGCTATTGCCATG 345
           | | |   ||    | ||   |       | |  |  | ||| ||    |   | || ||
Db     245 ATGGAGGTGGTGATGGTGATGGCGGCAGTGGCAGTGGTGATAGTGATGGAGGTGGCAATA 304

Qy     346 GTG---GCGGTCATTGTGGTGGTCGCAGTTGTCGGGAAAGGCGCGGCGGCGAAACTGGGT 402
           ||    | | | || |||||||| | | | || | | | |  | |  || ||   | |
Db     305 GTCATTGTGATGATGGTGGTGGTGGTAATGGTGGTGGAGGTGGTGATGGTGATGGTAGTG 364
```

```
Qy     403 AACGCGCTGAGCAAAATGATGGGCGAAACGATTAAGAAGTTGGT 446
           ||  ||   | | ||||||  |  | |||    | | ||||
Db     365 GTGGCAGTGGTGATAGTGATGGAGGTGATGATGGTCATGGTGGT 408


RESULT   2
AI238960
LOCUS       AI238960      527 bp    mRNA            EST       18-NOV-1998
DEFINITION  GH15115.5prime GH Drosophila melanogaster head pOT2 Drosophila
            melanogaster cDNA clone GH15115 5prime, mRNA sequence.
ACCESSION   AI238960
NID         g3833818
VERSION     AI238960.1  GI:3833818
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 527)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     On Aug 21, 1998 this sequence version replaced.

            Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Sequence is the complete cDNA Insert.
            Plate: 151  row: B  column: 3
            High quality sequence stop: 343.
FEATURES              Location/Qualifiers
     source           1. .527
                      /organism="Drosophila melanogaster"
                      /db_xref="taxon:7227"
                      /clone="GH15115"
                      /clone_lib="GH Drosophila melanogaster head pOT2"
                      /sex="male and female"
                      /dev_stage="adult"
                      /lab_host="DH5 - alpha"
                      /note="Organ: head; Vector: pOT2; Site_1: EcoRI; Site_2:
                      XhoI; Sized fractionated cDNAs were directly ligated into
                      pOT2.  Plasmid cDNA library."
BASE COUNT      131 a    108 c    166 g    122 t
ORIGIN
```

```
  Query Match             5.3%;  Score 46;  DB 43;  Length 527;
  Best Local Similarity   49.6%;  Pred. No. 0.021;
  Matches  118;  Conservative    0;  Mismatches  120;  Indels    0;  Gaps    0;

Qy     155 TTGTGAAGGCGGCGACGGGAGTGTCGTTTATTCAGCAGGCGCTAAACCCGATTATGGAGC 214
           ||  |  ||||| | |||  ||| ||    |     ||| | |   |   |  ||||||
Db      84 TTCCGCCGGCGGTGGCGGTGGTGGCGGAGGTGGCCAAGGAGGTTGGCAGAAGAATGGAGG 143

Qy     215 ATGTGCTGAAGCCGTTAATGGAGCTGATTGGCAAGGCGATTACCAAAGCGCTGGAAGGAT 274
           | | |  | ||  | |  |  ||  | ||||||  |     |    ||| ||| |
Db     144 AGGAGGAGGAGGAGGTGGTCAAGGAGGCTGGCAAAAGGGCGGCGGCGGCGGAGGAGGTGG 203

Qy     275 TAGGCGTCGATAAGAAAACGGCAGAGATGGCCGGCAGCATTGTTGGTGCGATTGTCGCCG 334
           |  | | ||    |    |||  ||| ||||  |||      || |   ||   || |  |
Db     204 CAAACATGGAGGTGGTGGCGGAGGAGGTGGCAAGCATGGAGGTGGAAACGGCGGTGGTGG 263

Qy     335 CTATTGCCATGGTGGCGGTCATTGTGGTGGTCGCAGTTGTCGGGAAAGGCGCGGCGGC 392
           |   |  ||| |||||||    | |||||  |  || |   | |  |  ||||||||
Db     264 AAAGCACGGTGGCGGCGGTGGAGGCGGTGGAGGTGGTGGAAAGCATGGAGGCGGCGGC 321


RESULT   3
```

```
D22789
LOCUS       D22789        686 bp    mRNA            EST       20-JUL-1998
DEFINITION  RICC1332A Rice callus Oryza sativa cDNA clone R, mRNA sequence.
ACCESSION   D22789
NID         g871921
VERSION     D22789.1  GI:871921
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1  (bases 1 to 686)
  AUTHORS   Sasaki,T. and Minobe,Y.
  TITLE     Rice cDNA from callus
  JOURNAL   Unpublished (1994)
COMMENT     On Dec 2, 1993 this sequence version replaced.

            Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program
            2-1-2 Kannondai,Tsukuba
            Ibaraki,
            Japan 305
            Tel: 0298-38-7441
            Fax: 0298-38-7468
            Email: tsasaki@abr.affrc.go.jp
            PROJECT ='RGP'.
FEATURES             Location/Qualifiers
     source          1. .686
                     /organism="Oryza sativa"
                     /strain="cultivar Nipponbare, sub_species Japonica"
                     /db_xref="taxon:4530"
                     /map="2"
                     /clone="R"
                     /clone_lib="Rice callus"
                     /note="Vector: pBluescript II SK+; Site_1: SalI; Site_2:
                     NotI; cDNA prepared from rice callus mRNAs by using
                     oligo(dT) as a primer and ligating to the SalI-NotI site
                     of pBluescript II SK+ phagemid.  "
BASE COUNT      125 a    132 c    246 g    183 t
ORIGIN
```

```
  Query Match             4.9%;  Score 42.8;  DB 23;  Length 686;
  Best Local Similarity   46.6%;  Pred. No. 0.16;
  Matches  137;  Conservative    0;  Mismatches  157;  Indels    0;  Gaps    0;

Qy     100 GGGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTG 159
           | || |    ||| | |||    || ||   | | ||| | |||||      |  || ||
Db     108 GTGGAGGTGGTGGTGGTGGTGAAGGAGGCGGTGGTGGTTACGGTGGATCGGGCTATGGTT 167

Qy     160 AAGGCGGCGACGGGAGTGTCGTTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTG 219
             ||    |      |||  |    |   | ||| ||       |  | |||   | |
Db     168 CTGGTTCTGGGTATGGTGAGGGAGGTGGCTCCGGCTCTGGCTATGGGTCTGGGCAAGGCT 227

Qy     220 CTGAAGCCGTTAATGGAGCTGATTGGCAAGGCGATTACCAAAGCGCTGGAAGGATTAGGC 279
           |||     | |   || | || | |    ||||  ||    || | |||  |   | |
Db     228 CTGGCTATGGTGCAGGTGTTGGTGGTGCTGGCGGGTATGGTAGTGGTGGCGGAGGTGGTG 287

Qy     280 GTCGATAAGAAAACGGCAGAGATGGCCGGCAGCATTGTTGGTGCGATTGTCGCCGCTATT 339
           || |  |||      ||    |  |||     |||  | |   || || ||  | |||||
Db     288 GTGGCCAAGGTGGTGGTGCTGGCGGCTACGGGCAAGGGTCTGGCTATGGTTCCGGCTATG 347

Qy     340 GCCATGGTGGCGGTCATTGTGGTGGTCGCAGTTGTCGGGAAAGGCGCGGCGGCG 393
           |    ||||  |||  |  |  |||| |  |||  | |     |  |||| ||||
Db     348 GGTCGGGTGCTGGTGGTGCTCATGGTGGTGGTTATGGAAGTGGTGGCGGTGGCG 401
```

```
RESULT   4
AI728199/c
LOCUS       AI728199      554 bp    mRNA            EST       11-JUN-1999
DEFINITION  BNLGHi10138 Six-day Cotton fiber Gossypium hirsutum cDNA 5' similar
            to (AL022537) putative protein [Arabidopsis thaliana], mRNA
            sequence.
ACCESSION   AI728199
NID         g5047051
VERSION     AI728199.1  GI:5047051
KEYWORDS    EST.
SOURCE      upland cotton.
  ORGANISM  Gossypium hirsutum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
REFERENCE   1  (bases 1 to 554)
  AUTHORS   Blewitt,M., Matz,E.C., Davy,D.F. and Burr,B.
  TITLE     ESTs from developing cotton fiber
  JOURNAL   Unpublished (1999)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3187982.

            Contact: Ben Burr
            Biology Department
            Brookhaven National Laboratory
            Upton, NY 11973, USA
            Tel: 516-344-3396
            Fax: 516-344-3407
            Email: burr@bnlux1.bnl.gov
            Seq primer: T3 Primer.
FEATURES             Location/Qualifiers
     source          1. .554
                     /organism="Gossypium hirsutum"
                     /cultivar="Acala Maxxa"
                     /db_xref="taxon:3635"
                     /clone_lib="Six-day Cotton fiber"
                     /tissue_type="immature fiber"
                     /dev_stage="Six days post anthesis"
                     /lab_host="XL1-Blue"
                     /note="Vector: pBluescript II KS+"
BASE COUNT      109 a    226 c     64 g    155 t
ORIGIN
```

```
  Query Match             4.8%;  Score 41.4;  DB 51;  Length 554;
  Best Local Similarity   49.5%;  Pred. No. 0.35;
  Matches  136;  Conservative    0;  Mismatches  136;  Indels    3;  Gaps    1;

Qy     101 GGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGA 160
           ||  ||  | |  || ||  |||  |||| | | || | ||| ||  || |  |  | ||
Db     351 GGAAGAAGCCGAGGCGGGAAGCGTCGGGATTGGGGGCATTGGCGGAGGAAGTGACCGAGA 292

Qy     161 AGGCGGCGACGGGAGTGTCGTTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGC 220
           || || |||||   || || |    | | ||  | | || |||| | |   | |   |
Db     291 CGGAGGAGACGGAGATGGCGATGTGACCGGAGTGGGTGAAGCCGAATTTCTCGAAGAGGA 232

Qy     221 TGAAGCCGTTAAT---GGAGCTGATTGGCAAGGCGATTACCAAAGCGCTGGAAGGATTAG 277
           | | |  |  |      |||| |||| |  | ||   | | |   ||| |||| |    |
Db     231 TCATGGGGCGAGAGTCGGAGGTGATGGTTAGGGTTTTGATCTCTGCGGTGGAGGAGGAAA 172

Qy     278 GCGTCGATAAGAAAACGGCAGAGATGGCCGGCAGCATTGTTGGTGCGATTGTCGCCGCTA 337
           || | ||  | |      || |  | | |    |  |   || ||| | | |
Db     171 TGAGGGAAAGGAGGAAGAGGAGGAGGATCTGTAAGGGTAGTTTCGCCATTTTGGGAGAAT 112

Qy     338 TTGCCATGGTGGCGGTCATTGTGGTGGTCGCAGTT 372
           || |   |||||      |  || |  || | ||||
Db     111 TTACGGCGGTGGTTACGACGGTCGGCGTGGTAGTT 77
```

```
RESULT   5
AI731721/c
LOCUS       AI731721      628 bp    mRNA            EST       11-JUN-1999
DEFINITION  BNLGHi10580 Six-day Cotton fiber Gossypium hirsutum cDNA 5' similar
            to (AL031032) extensin-like protein [Arabidopsis thaliana], mRNA
            sequence.
ACCESSION   AI731721
NID         g5050573
VERSION     AI731721.1  GI:5050573
```

```
KEYWORDS    EST.
SOURCE      upland cotton.
  ORGANISM  Gossypium hirsutum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
REFERENCE   1  (bases 1 to 628)
  AUTHORS   Blewitt,M., Matz,E.C., Davy,D.F. and Burr,B.
  TITLE     ESTs from developing cotton fiber
  JOURNAL   Unpublished (1999)
COMMENT     On Mar 10, 1998 this sequence version replaced gi:2948508.

            Contact: Ben Burr
            Biology Department
            Brookhaven National Laboratory
            Upton, NY 11973, USA
            Tel: 516-344-3396
            Fax: 516-344-3407
            Email: burr@bnlux1.bnl.gov
            Seq primer: T3 Primer.
FEATURES             Location/Qualifiers
     source          1. .628
                     /organism="Gossypium hirsutum"
                     /cultivar="Acala Maxxa"
                     /db_xref="taxon:3635"
                     /clone_lib="Six-day Cotton fiber"
                     /tissue_type="immature fiber"
                     /dev_stage="Six days post anthesis"
                     /lab_host="XL1-Blue"
                     /note="Vector: pBluescript II KS+"
BASE COUNT      144 a    306 c     44 g    134 t
ORIGIN
```

```
  Query Match             4.7%;  Score 41.2;  DB 51;  Length 628;
  Best Local Similarity   44.2%;  Pred. No. 0.42;
  Matches  169;  Conservative    0;  Mismatches  213;  Indels    0;  Gaps    0;

Qy      97 GGTGGGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATT 156
           || | |  |    ||| | ||| |      | | | | |||   ||||   |     |
Db     474 GGGGTGAAGGTGATGGAGGTGGTTTGTAATGAATTGGGGGTGGGGGTGATGGTGATGGTG 415

Qy     157 GTGAAGGCGGCGACGGGAGTGTCGTTTATTCAGCAGGCGCTAAACCCGATTATGGAGCAT 216
           |||  || || |  |   | ||    | | ||   || |||       | |  ||| | |
Db     414 GTGGGGGTGGGGGTGATGGCGTGTACTCTACACACGGTGCTGGTGGTGGTGGTGGTGGAG 355

Qy     217 GTGCTGAAGCCGTTAATGGAGCTGATTGGCAAGGCGATTACCAAAGCGCTGGAAGGATTA 276
           |||    |      |  |||| ||| |    || | | |   ||  |  ||| |
Db     354 GTGGCTCAATACATGGAGGAGGTGACGGTGGTGGTGGTGAGTGAACTGGAGGAGGTGGGG 295

Qy     277 GGCGTCGATAAGAAAACGGCAGAGATGGCCGGCAGCATTGTTGGTGCGATTGTCGCCGCT 336
             |  ||     ||| | | |||     || || | |     | | |  |  |
Db     294 AATAAACAGGAGGTGGTGGCGGTGGTGGAGATAAGTATGGATAAATTGGTGGAGGAGGTG 235

Qy     337 ATTGCCATGGTGGCGGTCATTGTGGTGGTCGCAGTTGTCGGGAAAGGCGCGGCGGCGAAA 396
           | ||   |||||| |||       ||    |  || |  | ||| || | ||||| |
Db     234 AATGTGGTGGTGGTGGTGGAGAATGTACCGGAGGTGGCGGTGAATGGTGTGGCGGTGGCG 175

Qy     397 CTGGGTAACGCGCTGAGCAAAATGATGGGCGAAACGATTAAGAAGTTGGTGCCTAACGTG 456
             | || |       |   || || |||   |    ||  |  | |  ||||    | | |
Db     174 GGGAGTTATAGTAGTAAGGAACTGGTGGAGGTGGGGAAAACAATGGAGGTGGTGGAGGGG 115

Qy     457 CTGAAACAGTTGGCGCAAAACG 478
           ||||   | ||| | |  | |
Db     114 GTGAATTTGGTGGTGGAGGAGG 93
```

```
RESULT   6
AI292820/c
LOCUS       AI292820      590 bp    mRNA            EST       01-DEC-1998
DEFINITION  GH15821.5prime GH Drosophila melanogaster head pOT2 Drosophila
            melanogaster cDNA clone GH15821 5prime, mRNA sequence.
ACCESSION   AI292820
NID         g3942227
VERSION     AI292820.1  GI:3942227
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 590)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     On Apr 14, 1993 this sequence version replaced gi:639346.

            Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            hit genomic sequence AL023893
            Plate: 158  row: B  column: 9
            High quality sequence stop: 433.
FEATURES             Location/Qualifiers
     source          1. .590
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="GH15821"
                     /clone_lib="GH Drosophila melanogaster head pOT2"
                     /sex="male and female"
                     /dev_stage="adult"
                     /lab_host="DH5 - alpha"
                     /note="Organ: head; Vector: pOT2; Site_1: EcoRI; Site_2:
                     XhoI; Sized fractionated cDNAs were directly ligated into
                     pOT2.  Plasmid cDNA library."
BASE COUNT      172 a    191 c    132 g     95 t
ORIGIN
```

```
  Query Match             4.7%;  Score 40.6;  DB 44;  Length 590;
  Best Local Similarity   49.3%;  Pred. No. 0.59;
  Matches  106;  Conservative    0;  Mismatches  109;  Indels    0;  Gaps    0;

Qy      42 GAAAGTCCTCGGCGCGCTGCTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTGG 101
           | | ||  |  |   || ||   ||  |||    || || |  | ||||  ||
Db     281 GCATGTGGTGCGGATGCGGCCCGCCCATGTTGTGGTGGTTGTTGTTGTTGTTATGAT 222

Qy     102 GGCGAGTCTGGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAA 161
           || |  | | | |  ||  ||  || |   |||  |  |||||||      |  | ||| |
Db     221 GGTGGTTGTTGGGGTGATTGTTGTTGTGGTTGTTGGGATGGTGGTGCTGCATGTGGTGCA 162

Qy     162 GGCGGCGACGGGAGTGTCGTTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCT 221
           |  |  || ||   ||  | |  |   |  |  |||    | ||| |||  ||    ||
Db     161 GCTGCGGATGGTGCTGCGGATGCTGGTGTGGATGCTGCGGCTGATGATGATGCGGCGGCA 102

Qy     222 GAAGCCGTTAATGGAGCTGATTGGCAAGGCGATTA 256
           | || | |  ||| | |  |||  | || |||||
Db     101 TATGCGGATTGTGGTGGTTGTTGTGATGGTGATTA 67
```

```
RESULT   7
AA749517/c
LOCUS       AA749517      487 bp    mRNA            EST       20-JAN-1998
DEFINITION  SWAMCAC06A12SK Brugia malayi adult male cDNA (SAW94NL-BmAM) Brugia
            malayi cDNA clone SWAMCAC06A12 5', mRNA sequence.
ACCESSION   AA749517
NID         g2796223
VERSION     AA749517.1  GI:2796223
KEYWORDS    EST.
SOURCE      Brugia malayi.
  ORGANISM  Brugia malayi
            Eukaryota; Metazoa; Nematoda; Secernentea; Spiruria; Spirurida;
```

```
            Filarioidea; Onchocercidae; Brugia.
REFERENCE   1  (bases 1 to 487)
  AUTHORS   Williams,S.A.
  TITLE     Genes expressed in adult males of Brugia malayi
  JOURNAL   Unpublished (1995)
COMMENT     On Jan 9, 1998 this sequence version replaced gi:693559.

            Contact: Steven A. Williams
            Molecular Parasitology
            Smith College Department of Biological Sciences
            Department of Biological Sciences, Clark Science Center, Smith
            College, Northampton, MA, 01063, USA
            Tel: 4135853826
            Fax: 4135853786
            Email: genome@smith.edu
            Seq primer: pBluescript SK.
FEATURES             Location/Qualifiers
     source          1. .487
                     /organism="Brugia malayi"
                     /strain="TRS Labs"
                     /db_xref="taxon:6279"
                     /clone="SWAMCAC06A12"
                     /clone_lib="Brugia malayi adult male cDNA (SAW94NL-BmAM)"
                     /lab_host="XL1-Blue MRF'"
                     /note="Vector: lambda UniZap XR; Site_1: EcoR I; Site_2:
                     Xho I; Lymphatic filarial nematode parasite of humans.
                     mRNA was prepared from adult males of Brugia malayi
                     isolated from jirds and converted to double stranded cDNA
                     using reverse transcriptase and oligo(dT) followed by
                     RNase H and DNApol I. The library had 4.6 x 10E6
                     independent recombinants and average insert size was 800
                     base pairs.  The library was constructed by Noelle Ling.
                     The library is available from Dr. S.A. Williams, email
                     genome@smith.edu."
BASE COUNT      138 a    175 c     71 g    102 t      1 others
ORIGIN
```

```
  Query Match             4.6%;  Score 40.2;  DB 38;  Length 487;
  Best Local Similarity   48.9%;  Pred. No. 0.7;
  Matches  108;  Conservative    0;  Mismatches  113;  Indels    0;  Gaps    0;

Qy      37 ATCGGGAAAGTCCTCGGCGCGCTGCTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACC 96
           || | ||  ||  | |  |  | |  |    | | ||  | ||| | |  | | ||
Db     256 ATGGAGATGGTGGTGGTGGTTCAGGCACTGGTGATGAGGATAGTGACGGTGGCTGTTCGG 197

Qy      97 GGTGGGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATT 156
           |    |||||    | |  |||| |  | ||   | | ||||||||||||  |  | | |
Db     196 GTATTGGCGATGACGACTGTGGCAGTAGCGGTGGTGGTGGTAATGGTGGTGGTGGTAGTG 137

Qy     157 GTGAAGGCGGCGACGGGAGTGTCGTTTATTCAGCAGGCGCTAAACCCGATTATGGAGCAT 216
           ||   || || | |||    |  | | ||   | ||  | |      |   || ||| |
Db     136 GTACCGGTGGTGGCGGTTCAGATGCTAATGATGGAGTTGGTGGCAATGGCGATAGAGAAA 77

Qy     217 GTGCTGAAGCCGTTAATGGAGCTGATTGGCAAGGCGATTAC 257
           | | ||  |  | |    | | |||||   ||| | | ||
Db      76 GGGATGCGGGAGGTGGGAGTGATGATTCAACAGGTGCTGAC 36
```

```
RESULT   8
AA390540/c
LOCUS       AA390540      650 bp    mRNA            EST       28-NOV-1998
DEFINITION  LD08780.5prime LD Drosophila melanogaster embryo BlueScript
            Drosophila melanogaster cDNA clone LD08780 5prime, mRNA sequence.
ACCESSION   AA390540
NID         g2043534
VERSION     AA390540.1  GI:2043534
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
```

```
REFERENCE   1  (bases 1 to 650)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1395282.

            Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 87  row: G  column: 8
            High quality sequence stop: 482.
FEATURES             Location/Qualifiers
     source          1. .650
                     /organism="Drosophila melanogaster"
                     /db_xref="BDGP_EST:BDcln008107"
                     /db_xref="taxon:7227"
                     /clone="LD08780"
                     /clone_lib="LD Drosophila melanogaster embryo BlueScript"
                     /sex="male and female"
                     /dev_stage="0 to 24 hours mixed stage embryonic"
                     /lab_host="SOLR"
                     /note="Organ: embryo; Vector: BlueScript SK; Site_1:
                     EcoRI; Site_2: XhoI; Constructed using Stratagene ZAP-cDNA
                     Synthesis kit. Oligo dT-primed and directionally cloned at
                     EcoRI and XhoI in BlueScript SK(+/-)"
BASE COUNT      182 a    216 c    158 g     94 t
ORIGIN
```

```
  Query Match             4.5%;  Score 39;  DB 33;  Length 650;
  Best Local Similarity   48.8%;  Pred. No. 1.6;
  Matches  105;  Conservative    0;  Mismatches  110;  Indels    0;  Gaps    0;

Qy      42 GAAAGTCCTCGGCGCGCTGCTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTGG 101
           | | ||  |  |   || |    ||  |||    || || |  | |||| ||
Db     579 GCATGTGGTGCGGATGCGGGCCGCCCATGTTGTGGTGGTTGTTGTTGTTGTTGTTATGAT 520

Qy     102 GGCGAGTCTGGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAA 161
           || |  | | | | ||  ||  || |   |||  |  |||||||      |  | |||
Db     519 GGTGGTTGTTGGGGTGATTGTTGTTGTGGTTGTTGGGATGGTGGTGCTGCATGTGGTGCA 460

Qy     162 GGCGGCGACGGGAGTGTCGTTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCT 221
           |  |  || ||   ||  | | |   |  |  |||    | ||| |||  ||    ||
Db     459 GCTGCGGATGGTGCTGCGGATGCTGGTGTGGATGCTGCGGCTGATGATGATGCGGCGGCA 400

Qy     222 GAAGCCGTTAATGGAGCTGATTGGCAAGGCGATTA 256
           | || | |  ||| | |  |||  | || |||||
Db     399 TATGCGGATTGTGGTGGTTGTTGTGATGGTGATTA 365
```

```
RESULT   9
AI292805/c
LOCUS       AI292805      616 bp    mRNA            EST       01-DEC-1998
DEFINITION  GH15806.5prime GH Drosophila melanogaster head pOT2 Drosophila
            melanogaster cDNA clone GH15806 5prime, mRNA sequence.
ACCESSION   AI292805
NID         g3942212
VERSION     AI292805.1  GI:3942212
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 616)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
```

```
COMMENT     On May 1, 1998 this sequence version replaced gi:2761092.

            Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            hit genomic sequence AL023893
            Plate: 158  row: A  column: 6
            High quality sequence stop: 544.
FEATURES             Location/Qualifiers
     source          1. .616
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="GH15806"
                     /clone_lib="GH Drosophila melanogaster head pOT2"
                     /sex="male and female"
                     /dev_stage="adult"
                     /lab_host="DH5 - alpha"
                     /note="Organ: head; Vector: pOT2; Site_1: EcoRI; Site_2:
                     XhoI; Sized fractionated cDNAs were directly ligated into
                     pOT2.  Plasmid cDNA library."
BASE COUNT      178 a    202 c    138 g     98 t
ORIGIN


  Query Match             4.5%;  Score 39;  DB 44;  Length 616;
  Best Local Similarity   48.8%;  Pred. No. 1.6;
  Matches  105;  Conservative    0;  Mismatches  110;  Indels    0;  Gaps    0;

Qy      42 GAAAGTCCTCGGCGCGCTGCTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTGG 101
           | | ||  |  |   || |    ||  |||    || || |  | |||| ||
Db     281 GCATGTGGTGCGGATGCGGGCCGCCCATGTTGTGGTGGTTGTTGTTGTTGTTGTTATGAT 222

Qy     102 GGCGAGTCTGGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAA 161
           || |  | | | ||  ||  || |   |||  |  |||||||      |  | ||| |
Db     221 GGTGGTTGTTGGGGTGATTGTTGTTGTGGTTGTTGGGATGGTGGTGCTGCATGTGGTGCA 162

Qy     162 GGCGGCGACGGGAGTGTCGTTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCT 221
           |  |  || ||   ||  | |  |   |  |  |||    | ||| |||  ||    ||
Db     161 GCTGCGGATGGTGCTGCGGATGCTGGTGTGGATGCTGCGGCTGATGATGATGCGGCGGCA 102

Qy     222 GAAGCCGTTAATGGAGCTGATTGGCAAGGCGATTA 256
           | || | |  ||| | |  |||  | || |||||
Db     101 TATGCGGATTGTGGTGGTTGTTGTGATGGTGATTA 67


RESULT  10
AI292818/c
LOCUS       AI292818      468 bp    mRNA            EST       01-DEC-1998
DEFINITION  GH15819.5prime GH Drosophila melanogaster head pOT2 Drosophila
            melanogaster cDNA clone GH15819 5prime, mRNA sequence.
ACCESSION   AI292818
NID         g3942225
VERSION     AI292818.1  GI:3942225
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 468)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     On Apr 14, 1993 this sequence version replaced gi:639324.

            Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            hit genomic sequence AL023893
            Plate: 158  row: B  column: 7
            High quality sequence stop: 433.
FEATURES             Location/Qualifiers
     source          1. .468
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="GH15819"
                     /clone_lib="GH Drosophila melanogaster head pOT2"
                     /sex="male and female"
                     /dev_stage="adult"
                     /lab_host="DH5 - alpha"
                     /note="Organ: head; Vector: pOT2; Site_1: EcoRI; Site_2:
                     XhoI; Sized fractionated cDNAs were directly ligated into
                     pOT2.  Plasmid cDNA library."
BASE COUNT      139 a    161 c    109 g     59 t
ORIGIN


  Query Match             4.5%;  Score 39;  DB 44;  Length 468;
  Best Local Similarity   48.8%;  Pred. No. 1.4;
  Matches  105;  Conservative    0;  Mismatches  110;  Indels    0;  Gaps    0;

Qy      42 GAAAGTCCTCGGCGCGCTGCTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTGG 101
           | | ||  |  |   || |    ||  |||    || || |  | |||| ||
Db     281 GCATGTGGTGCGGATGCGGGCCGCCCATGTTGTGGTGGTTGTTGTTGTTGTTGTTATGAT 222

Qy     102 GGCGAGTCTGGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAA 161
           || |  | | | ||  ||  || |   |||  |  |||||||      |  | ||| |
Db     221 GGTGGTTGTTGGGGTGATTGTTGTTGTGGTTGTTGGGATGGTGGTGCTGCATGTGGTGCA 162

Qy     162 GGCGGCGACGGGAGTGTCGTTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCT 221
           |  |  || ||   ||  | |  |   |  |  |||    | ||| |||  ||    ||
Db     161 GCTGCGGATGGTGCTGCGGATGCTGGTGTGGATGCTGCGGCTGATGATGATGCGGCGGCA 102

Qy     222 GAAGCCGTTAATGGAGCTGATTGGCAAGGCGATTA 256
           | || | |  ||| | |  |||  | || |||||
Db     101 TATGCGGATTGTGGTGGTTGTTGTGATGGTGATTA 67


RESULT  11
AA441313/c
LOCUS       AA441313      531 bp    mRNA            EST       27-NOV-1998
DEFINITION  LD16203.5prime LD Drosophila melanogaster embryo BlueScript
            Drosophila melanogaster cDNA clone LD16203 5prime, mRNA sequence.
ACCESSION   AA441313
NID         g2794159
VERSION     AA441313.1  GI:2794159
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 531)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2153191.

            Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 162  row: A  column: 3
            High quality sequence stop: 432.
FEATURES             Location/Qualifiers
     source          1. .531
```

```
                     /organism="Drosophila melanogaster"
                     /db_xref="BDGP_EST:BDcln015422"
                     /db_xref="taxon:7227"
                     /clone="LD16203"
                     /clone_lib="LD Drosophila melanogaster embryo BlueScript"
                     /sex="male and female"
                     /dev_stage="0 to 24 hours mixed stage embryonic"
                     /lab_host="SOLR"
                     /note="Organ: embryo; Vector: BlueScript SK; Site_1:
                     EcoRI; Site_2: XhoI; Constructed using Stratagene ZAP-cDNA
                     Synthesis kit. Oligo dT-primed and directionally cloned at
                     EcoRI and XhoI in BlueScript SK(+/-)"
BASE COUNT      149 a    175 c    139 g     67 t      1 others
ORIGIN


  Query Match             4.4%;  Score 38.6;  DB 38;  Length 531;
  Best Local Similarity   49.7%;  Pred. No. 1.9;
  Matches   98;  Conservative    0;  Mismatches   99;  Indels    0;  Gaps    0;

Qy      60 GCTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGC 119
           ||   ||  |||    || || |  | |||| ||        || |  | | | | ||
Db     343 GCCGCCCCATGTTGTGGTGGTTGTTGTTGTTGTTATGATGGTGGTTGTTGGGGTGAT 284

Qy     120 TGCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTC 179
           ||  || |   |||  |  |||||||       |  | ||| ||  |  || ||   ||
Db     283 TGTTGTTGTGGTTGTTGGGATGGTGGTGCTGCATGTGGTGCAGCTGCGGATGGTGCTGCG 224

Qy     180 GTTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCT 239
           | |  |   |  |  |||    | ||| |||  ||    ||  | || | |  ||| | |
Db     223 GATGCTGGTGTGGATGCTGCGGCTGATGATGATGCGGCGGCATATGCGGATTGTGGTGGT 164

Qy     240 GATTGGCAAGGCGATTA 256
             |||  | ||  |||||
Db     163 TGTTGTGATGGTGATTA 147


RESULT  12
H76570
LOCUS       H76570         564 bp    mRNA            EST       05-JAN-1998
DEFINITION  18275 Lambda-PRL2 Arabidopsis thaliana cDNA clone 198H4T7, mRNA
            sequence.
ACCESSION   H76570
NID         g1053821
VERSION     H76570.1  GI:1053821
KEYWORDS    EST.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
            Arabidopsis.
REFERENCE   1  (bases 1 to 564)
  AUTHORS   Newman,T., deBruijn,F.J., Green,P., Keegstra,K., Kende,H.,
            McIntosh,L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomashow,M.,
            Retzel,E. and Somerville,C.
  TITLE     Genes galore: a summary of methods for accessing results from
            large-scale partial sequencing of anonymous Arabidopsis cDNA clones
  JOURNAL   Plant Physiol. 106, 1241-1255 (1994)
  MEDLINE   95148729
COMMENT     On May 9, 1995 this sequence version replaced gi:802694.

            Contact: Thomas Newman
            MSU-DOE Plant Research Laboratory
            Michigan State University
            MSU-DOE-PRL, Michigan State University,Plant Biology Bldg.,E.
            Lansing,Mi
            Tel: 517-353-0854
            Fax: 517-353-9168
            Email: 22313tcn@ibm.cl.msu.edu
            Seq primer: 5_prime.
FEATURES             Location/Qualifiers
     source          1. .564
                     /organism="Arabidopsis thaliana"
                     /strain="var columbia"
                     /db_xref="taxon:3702"
                     /clone="198H4T7"
                     /clone_lib="Lambda-PRL2"
                     /note="Vector: lambda Zip-Lox; Site_1: Sal; Site_2: Not;
                     Lambda PRL2 is a cDNA library derived from equal
                     quantities of 4 pools of mRNA.  The mRNA sources were 1) 7
                     day germinated etiolated seedlings; 2) tissue culture
                     grown roots; 3) staged plants half with 24 hour light
                     cycle, half on 16 hr light, 8 hour dark- rosettes;  4)
                     same plants as 3 but aerial tissue (stems, flowers and
                     siliques.  The vector is BRL's lambda Zip-Lox.  The cDNA
                     inserts were directionally cloned with Sal-Not arms using
                     oligo dT primed cDNA.  "
BASE COUNT      127 a     75 c    202 g    138 t     22 others
ORIGIN


  Query Match             4.4%;  Score 38;  DB 24;  Length 564;
  Best Local Similarity   48.4%;  Pred. No. 2.9;
  Matches  132;  Conservative    0;  Mismatches  135;  Indels    6;  Gaps    1;

Qy     100 GGGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTG 159
           ||||  |    || | |||  |||||||  | | | ||   ||| ||  || |   | | |
Db      68 GGGGACATGGAGGTGGTGGAGGCGGTGGAAATGGTGGGGGTGGAGGAGGAGGTTCTGGCG 127

Qy     160 AAGGCGGCGACGGGAGTGTCGTTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTG 219
           |||| || |      |||  | |||   ||||||  | |  | |      | |||    |||
Db     128 AAGGAGGAGCTCATGGTGGTGGTTACGGAGCAGGAGGTGGAGCTNGAGAGGGATACGGTG 187

Qy     220 CTGAAGCCGTTAATGGAGCTGATTGGCAAGGCGATTACCAAAGCGCTGGAAGGATTAGGC 279
             ||  ||      ||||||||  | | || || | |    | | |  ||  | |
Db     188 GTGGTGC------TGGAGCTGGTGGACATGGTGGTGGAGGAGGAGGCGGTGGTGGTTCCN 241

Qy     280 GTCGATAAGAAAACGGCAGAGATGGCCGGCAGCATTGTTGGTGCGATTGTCGCCGCTATT 339
           |  |   || |   ||  | |  ||    |    |  ||   |  |  ||| | |
Db     242 GAGGTGGAGGAGGAGGAGGTGGAGGTTATCCTGCTGCTTCTGGATACGGTCACGGTNGGG 301

Qy     340 GCCATGGTGGCGGTCATTGTGGTGGTCGCAGTT 372
           |   | ||||  |  |  |  | ||||||  ||
Db     302 GTGCTNGTGGTNGAGAAGGAAGCGGTCGCTATT 334


RESULT  13
AI067558/c
LOCUS       AI067558       619 bp    mRNA            EST       03-AUG-1998
DEFINITION  EST209240 Schistosoma mansoni, Phil LoVerde/Joe Merrick Schistosoma
            mansoni cDNA clone SMNBD47 5' end similar to eggshell protein
            (HT:62583), mRNA sequence.
ACCESSION   AI067558
NID         g3385525
VERSION     AI067558.1  GI:3385525
KEYWORDS    EST.
SOURCE      Schistosoma mansoni.
  ORGANISM  Schistosoma mansoni
            Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
            Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.
REFERENCE   1  (bases 1 to 619)
  AUTHORS   Merrick,J.M., Osman,A., LoVerde,P.T., Chandra,I., Glodek,A.,
            Fraser,C.M. and Lee,N.H.
  TITLE     Generation of a Schistosoma sp. EST (SMEST) Catalog & Schistosome
            Gene Index
  JOURNAL   Unpublished (1998)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1397399.

            Contact: Merrick, J.
            State University of New York at Buffalo
            Department of Microbiology, School of Medicine and Biomedical
            Sciences, Buffalo, NY 14214,USA
            Tel: (716)-829-2158
            Fax: (716)-829-2177
            Email: merrick@acsu.buffalo.edu
```

```
          Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .619
                     /organism="Schistosoma mansoni"
                     /db_xref="taxon:6183"
                     /clone="SMNBD47"
                     /clone_lib="Schistosoma mansoni, Phil LoVerde/Joe Merrick"
                     /note="Vector: pBluescript SK; Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT      169 a    199 c     93 g    158 t
ORIGIN


  Query Match             4.3%;  Score 37.8;  DB 41;  Length 619;
  Best Local Similarity   50.8%;  Pred. No. 3.4;
  Matches   90;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;

Qy     295 GCAGAGATGGCCGGCAGCATTGTTGGTGCGATTGTCGCCGCTATTGCCATGGTGGCGGTC 354
           || |  |||||          | | |  | |  |||    |   |  |  ||| |||||
Db     603 GCGGTTATGGCGATAGTGGATATGGTGGAGGCTGTACTGGTGGTGACTGTGGCGGCGGCT 544

Qy     355 ATTGTGGTGGTCGCAGTTGTCGGGAAAGGCGCGGCGGCGAAACTGGGTAACGCGCTGAGC 414
           || |||| |||  |||| || | ||  |  |       ||    ||| ||  | | |||
Db     543 ATGGTGGAGGTTACAGTGGTGGAGATTGTGGTAATTACGGTGGTGGCTATGGTGGTGATT 484

Qy     415 AAAATGATGGGCGAAACGATTAAGAAGTTGGTGCCTAACGTGCTGAAACAGTTGGCG 471
             |||| |||       | | |  | | ||||| |||  ||| ||  |  | ||| |
Db     483 GCAATGGTGGAGATTGTGGTAATTACGGTGGTGGCTATGGTGGTGGGAATGGTGGTG 427


RESULT  14
AU033900
LOCUS       AU033900      420 bp    mRNA            EST       28-APR-1999
DEFINITION  AU033900 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium
            discoideum cDNA clone SLB585, mRNA sequence.
ACCESSION   AU033900
NID         g3799324
VERSION     AU033900.1  GI:3799324
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Dictyostelida; Dictyostelium.
REFERENCE   1  (bases 1 to 420)
  AUTHORS   Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
            Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
            Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
  TITLE     The Dictyostelium developmental cDNA project: generation and
            analysis of expressed sequence tags from the first-finger stage of
            development
  JOURNAL   DNA Res. 5 (6), 335-340 (1998)
  MEDLINE   99156227
COMMENT     On Jan 14, 1998 this sequence version replaced gi:1797031.

            Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan
            Email: d402hu@sakura.cc.tsukuba.ac.jp
            PROJECT = 'Dictyostelium discoideum cDNA project in Japan'
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .420
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="SLB585"
                     /clone_lib="Dictyostelium discoideum SL (H.Urushihara)"
                     /dev_stage="slug"
BASE COUNT      121 a     44 c    140 g    114 t      1 others
ORIGIN


  Query Match             4.3%;  Score 37.8;  DB 43;  Length 420;
  Best Local Similarity   51.2%;  Pred. No. 2.9;
  Matches   87;  Conservative    0;  Mismatches   83;  Indels    0;  Gaps    0;

Qy     210 GGAGCATGTGCTGAAGCCGTTAATGGAGCTGATTGGCAAGGCGATTACCAAAGCGCTGGA 269
           || |  | || ||        | |||| | |||| |  |||| | |||| | || | |||
Db       6 GGTGGTTATGGTGGTAATCGTGATGGTGGTGATNGAGAAGGTGGTTACAACAGAGGTGGC 65

Qy     270 AGGATTAGGCGTCGATAAGAAAACGGCAGAGATGGCCGGCAGCATTGTTGGTGCGATTGT 329
             | |   |||  | |      |  ||||| ||  |  |     || | |  | | ||
Db      66 TATGGTGGTAATCGTGATGGTGGTGATAGAGAAGGTGGCTACAGAGGTAGTTATGGTGGT 125

Qy     330 CGCCGCTATTGCCATGGTGGCGGTCATTGTGGTGGTCGCAGTTGTCGGGA 379
             || |  || |  | |||||     |  | ||||||    ||  ||| ||
Db     126 GGCGGTGATAGAGAAGGTGGTTACAACAGAGGTGGTTATGGTAATCGTGA 175


RESULT  15
AI389106/c
LOCUS       AI389106      676 bp    mRNA            EST       28-JAN-1999
DEFINITION  GH20192.5prime GH Drosophila melanogaster head pOT2 Drosophila
            melanogaster cDNA clone GH20192 5prime, mRNA sequence.
ACCESSION   AI389106
NID         g4203117
VERSION     AI389106.1  GI:4203117
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 676)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     On Nov 2, 1998 this sequence version replaced gi:3831105.

            Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 201  row: H  column: 8
            High quality sequence stop: 542.
FEATURES             Location/Qualifiers
     source          1. .676
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="GH20192"
                     /clone_lib="GH Drosophila melanogaster head pOT2"
                     /sex="male and female"
                     /dev_stage="adult"
                     /lab_host="DH5 - alpha"
                     /note="Organ: head; Vector: pOT2; Site_1: EcoRI; Site_2:
                     XhoI; Sized fractionated cDNAs were directly ligated into
                     pOT2.  Plasmid cDNA library."
BASE COUNT      248 a    246 c    100 g     82 t
ORIGIN


  Query Match             4.3%;  Score 37.8;  DB 45;  Length 676;
  Best Local Similarity   44.9%;  Pred. No. 3.5;
  Matches  144;  Conservative    0;  Mismatches  177;  Indels    0;  Gaps    0;

Qy      68 TTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCTGCGGTGG 127
           | |||  |||||| |  | |||| ||   |  |  | |  | | |   | |  | |||||
Db     605 TCGTCGTCGTTGTTGTGGTTGTTGTTGTTGTCGTCGTGGTTGTTGTTGTCGTCGTGGTGG 546

Qy     128 GACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCGTTTATTC 187
             |||  ||  |||| | ||  | | | ||    | || |  ||  ||        | |
Db     545 TTGTTGTTGTCGTGGTTGTCGTAGTAGTAGTTGTTGTGGTGGTGGTCGTCGTTGGAGTAC 486
```

```
Qy    188 AGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTGATTGGCA 247
          |    |   |       | |  | |    |||  |       | |  | |   |  | |
Db    485 ACGTTGTCGTCGTGGTGGTGGTTGTTGTTGTCGTCGTCGTGGTTGTCGTAGTAGTAGTTG 426

Qy    248 AGGCGATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAGATGGCCG 307
            |   |   |   |  | |  |  | |  || | |        || | |  | |  |
Db    425 TTGTCGTCGTCGTCGTTGTTGTTGTTGGTGGTGGTTGTTGTCGGGGAACATGTCGTTG 366

Qy    308 GCAGCATTGTTGGTGCGATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTGGTGGTCG 367
              | | ||||  | | | || |  | | | |   | |||| |||  |||| || || |
Db    365 TTGTCGTAGTTGTGGTGGTAGTTGTTGTTGTCGTAGTTGTGGTGGTTGTTGTTGTTGTGG 306

Qy    368 CAGTTGTCGGGAAAGGCGCGG 388
           |||||| |    || || ||
Db    305 TAGTTGTTGTTGTAGTCGTGG 285


Search completed: October  5, 1999, 17:13:44
Job time: 8667 sec
```

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 1999, 14:49:20 ; Search time 140.41 Seconds
(without alignments)
644.667 Million cell updates/sec

Title: US-09-068-804-1
Perfect score: 870
Sequence: 1 CGCAAAGCCGAGGAAACGAA.....TGCGCCAGAGTCGCGCATAA 870

Scoring table: IDENTITY_NUC

Searched: 192659 seqs, 52021692 residues

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/5C_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/5D_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	54	6.2	7218	2	US-08-232-463-14	Sequence 14, Appl
2	45.2	5.2	1505	1	US-07-915-246-1	Sequence 1, Appli
c 3	38.2	4.4	10366	2	US-08-246-982A-5	Sequence 5, Appli
c 4	38.2	4.4	10366	2	US-08-453-265-5	Sequence 5, Appli
c 5	38.2	4.4	10348	3	US-08-457-273B-41	Sequence 41, Appl
c 6	35.8	4.1	1295	2	US-08-433-854-3	Sequence 3, Appli
7	35.8	4.1	2793	2	US-08-209-747-1	Sequence 1, Appli
c 8	35.8	4.1	1295	2	US-08-174-745A-3	Sequence 3, Appli
9	35.8	4.1	2793	2	US-08-458-298-1	Sequence 1, Appli
c 10	35.8	4.1	1295	3	US-08-195-947-3	Sequence 3, Appli
c 11	35.8	4.1	1295	3	US-08-433-885-3	Sequence 3, Appli
12	34.8	4.0	6530	4	US-08-146-930-1	Sequence 1, Appli
13	34.8	4.0	6530	5	PCT-US93-03993-1	Sequence 1, Appli
14	34.2	3.9	10747	4	US-08-147-777-1	Sequence 1, Appli
15	34.2	3.9	24979	4	US-08-147-777-3	Sequence 3, Appli
16	34.2	3.9	10747	5	PCT-US93-03985-1	Sequence 1, Appli
17	34.2	3.9	24979	5	PCT-US93-03985-3	Sequence 3, Appli
18	34	3.9	2338	2	US-08-425-069-1	Sequence 1, Appli
19	33.4	3.8	3319	1	US-08-006-676B-2	Sequence 2, Appli
20	33.4	3.8	3319	2	US-08-282-845-1	Sequence 1, Appli
21	33.4	3.8	3319	4	US-08-428-414A-4	Sequence 4, Appli
22	33.4	3.8	3319	5	PCT-US94-00324-2	Sequence 2, Appli
23	32.4	3.7	564	3	US-08-775-978-2	Sequence 2, Appli
c 24	32.2	3.7	9757	1	US-08-093-453B-1	Sequence 1, Appli
c 25	32.2	3.7	9759	2	US-08-459-041A-1	Sequence 1, Appli
26	31.8	3.7	894	3	US-08-454-720A-32	Sequence 32, Appl
27	31.6	3.6	2814	1	US-07-781-034-1	Sequence 1, Appli
28	31.6	3.6	1288	2	US-08-440-856A-9	Sequence 9, Appli
29	31.6	3.6	4835	2	US-08-386-495-9	Sequence 9, Appli
30	31.6	3.6	2814	5	PCT-US92-08328-1	Sequence 1, Appli
31	31.6	3.6	4835	5	PCT-US96-02331-9	Sequence 9, Appli
32	31.6	3.6	4255	5	PCT-US96-02331-14	Sequence 14, Appl
33	31.2	3.6	8438	1	US-07-945-283-1	Sequence 1, Appli
34	31	3.6	1236	2	US-08-440-856A-1	Sequence 1, Appli
c 35	31	3.6	1015	3	US-08-121-436A-1	Sequence 1, Appli
36	31	3.6	7286	5	PCT-US95-11684-1	Sequence 1, Appli
c 37	30.8	3.5	2830	1	US-07-882-292-1	Sequence 1, Appli
38	30.8	3.5	1479	1	US-08-476-008-68	Sequence 68, Appl
39	30.8	3.5	1479	1	US-08-306-063-68	Sequence 68, Appl
40	30.8	3.5	7362	3	US-08-455-543A-7	Sequence 7, Appli
41	30.8	3.5	7175	3	US-08-455-543A-8	Sequence 8, Appli
42	30.8	3.5	1479	3	US-08-833-485-68	Sequence 68, Appl
43	30.8	3.5	7362	3	US-08-193-078B-7	Sequence 7, Appli
44	30.8	3.5	7175	3	US-08-193-078B-8	Sequence 8, Appli
45	30.8	3.5	7362	3	US-08-223-305C-7	Sequence 7, Appli

ALIGNMENTS

```
RESULT   1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14


  Query Match             6.2%;  Score 54;  DB 2;  Length 7218;
  Best Local Similarity   4.1%;  Pred. No. 4e-06;
  Matches   15;  Conservative  209;  Mismatches  144;  Indels    0;  Gaps    0;

Qy     259 AAAGCGCTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAGATGGCCGGCAGCATTGTT 318
           |||| | | |||| ||| ||    :: ::::::: :: ::::: ::  :: :: :  :
Db    1457 AAAGAGATAGAAGAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1398

Qy     319 GGTGCGATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTGGTGGTCGCAGTTGTCGGG 378
           :: : ::  :  :  :  :  :  : :: :: ::  :  : :: ::  : ::  :  :::
```

```
Db   1397 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1338

Qy    379 AAAGGCGCGGCGGCGAAACTGGGTAACGCGCTGAGCAAAATGATGGGCGAAACGATTAAG 438
          ::::: : :: :: :::: ::: :: : :  ::: :::: :: ::: :::: ::  :::
Db   1337 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1278

Qy    439 AAGTTGGTGCCTAACGTGCTGAAACAGTTGGCGCAAAACGGCAGCAAACTCTTTACCCAG 498
          :::  :: :   :: : :  :::: ::  :: : :::: :: :: :::     :   ::
Db   1277 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1218

Qy    499 GGGATGCAACGTATTACTAGCGGTCTGGGTAATGTGGGTAGCAAGATGGGCCTGCAAACG 558
          :::: : :: : :  :  :: ::  ::: :: : ::: :: :::: :::   : ::: :
Db   1217 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1158

Qy    559 AATGCCTTAAGTAAAGAGCTGGTAGGTAATACCCTAAATAAAGTGGCGTTGGGCATGGAA 618
          :: :    ::: :::::  :: ::: :: :    ::: :::: :: :  ::: : ::::
Db   1157 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1098

Qy    619 GTCACGAA 626
          :  : :::
Db   1097 RRRRRRRR 1090
```

```
RESULT   2
US-07-915-246-1
; Sequence 1, Application US/07915246
; Patent No. 5401836
;  GENERAL INFORMATION:
;     APPLICANT:  Baszczynski, Chris L.
;     APPLICANT:  Fallis, Lynne
;     APPLICANT:  Bellmare, Guy
;     APPLICANT:  Boivin, Rodolphe
;     TITLE OF INVENTION:  A BRASSICA REGULATORY SEQUENCE FOR
;     TITLE OF INVENTION:  ROOT-SPECIFIC OR ROOT-ABUNDANT GENE EXPRESSION
;     NUMBER OF SEQUENCES:  5
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  McAndrews, Held, and Malloy
;       STREET:  500 W. Madison St. 34th Floor
;       CITY:  Chicago
;       STATE:  Illinois
;       COUNTRY:  USA
;       ZIP:  60661
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/915,246
;       FILING DATE:  19920716
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Pochopien, Donald J.
;       REGISTRATION NUMBER:  32,167
;       REFERENCE/DOCKET NUMBER:  91 P 1125
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  312 707-8889
;       TELEFAX:  312 707-9155
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1505 base pairs
;       TYPE:  NUCLEIC ACID
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  Brassica napus
;       STRAIN:  Westar
;       DEVELOPMENTAL STAGE:  Somatic
;       TISSUE TYPE:  Root
US-07-915-246-1
```

```
  Query Match             5.2%;  Score 45.2;  DB 1;  Length 1505;
  Best Local Similarity   47.2%;  Pred. No. 0.00067;
  Matches  137;  Conservative    0;  Mismatches  153;  Indels    0;  Gaps    0;

Qy    111 GGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGAC 170
          || | |||  ||||||| | | | ||   ||| ||  |  ||  |    | || || |
Db   1047 GGTGGTGGAGGCGGTGGAAATGGAGGCGGTGGAGGAGGTGGAGGTGCACACGGTGGTGGA 1106

Qy    171 GGGAGTGTCGTTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTT 230
            |||  |   |    || || |||  |   |  |||||||   | | || ||  |
Db   1107 TACGGTGGTGGAGAAGGCGCTGGTGCTGGAGGAGGATATGGAGGTGGCGGTGCAGGTGGA 1166

Qy    231 AATGGAGCTGATTGGCAAGGCGATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAAGAA 290
          |||||| || | |    || |   |   | ||| |||| |     |   | | |    |
Db   1167 CATGGAGGTGGTGGAGGCGGGGGAAAAGGAGGCGGTGGAGGAGGAGGTTCTGGCGCCGGT 1226

Qy    291 AACGGCAGAGATGGCCGGCAGCATTGTTGGTGCGATTGTCGCCGCTATTGCCATGGTGGC 350
          |     | |||  |  |   |   |  | |   | |     | |  ||||||
Db   1227 GGAGCTCACGGTGGTGGTTATGGTGCCGGAGGTGGAGCTGGAGAGGGATACGGTGGTGGT 1286

Qy    351 GGTCATTGTGGTGGTCGCAGTTGTCGGGAAAGGCGCGGCGGCGAAACTGG 400
          |||      || || | | || || | | | |  | || || | | ||||
Db   1287 GGTGGAGAAGGAGGACACGGTGGTGGAGGAGGCGGTGGTGGTGGAGCTGG 1336
```

```
RESULT   3
US-08-246-982A-5/c
; Sequence 5, Application US/08246982A
; Patent No. 5686288
;  GENERAL INFORMATION:
;     APPLICANT:  MacDonald, Marcy E.
;     APPLICANT:  Ambrose, Christine M.
;     APPLICANT:  Duyao, Mabel P.
;     APPLICANT:  Gusella, James F.
;     TITLE OF INVENTION:  Huntingtin DNA, Protein And Uses Thereof
;     NUMBER OF SEQUENCES:  25
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Sterne, Kessler, Goldstein & Fox
;       STREET:  1100 New York Avenue
;       CITY:  Washington
;       STATE:  D.C.
;       COUNTRY:  U.S.A.
;       ZIP:  20005
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/246,982A
;       FILING DATE:  May 20, 1994
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Goldstein, Jorge, A.
;       REGISTRATION NUMBER:  29,021
;       REFERENCE/DOCKET NUMBER:  0609.3880002
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (202) 371-2600
;       TELEFAX:  (202) 371-2540
;   INFORMATION FOR SEQ ID NO:  5:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  10366 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  316..9748
US-08-246-982A-5
```

```
  Query Match             4.4%;  Score 38.2;  DB 2;  Length 10366;
  Best Local Similarity   51.6%;  Pred. No. 0.15;
  Matches  111;  Conservative    0;  Mismatches  103;  Indels    1;  Gaps    1;

Qy      74 GCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTG 133
           ||   ||||| || |      | |  ||||||  |  |||   ||| || | ||   |
Db     552 GCCGGGTGGCGGCGGGGGCGGCGGCGGGGGCGGCTGCGGCTGAGGCAGCAGCGGCTGTGC 493

Qy     134 CGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCGTTTAT-TCAGCAG 192
           | |    || ||| || |||  || | ||||||||  |||  | | ||    |  | || |
Db     492 CTGCGGCGGCGGCTGAGGAAGCTGAGGAGGCGGCGGCGGCGGCGGCGGTGGCGGCTG 433

Qy     193 GCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTGATTGGCAAGGCG 252
             |||    | | |  ||  ||   |||||  || | |  ||  |||| |      |  |
Db     432 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 373

Qy     253 ATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAA 287
           |    |||  || || ||| | |  | |  | |
Db     372 CTGCTGGAAGGACTTGAGGGACTCGAAGGCCTTCA 338


RESULT   4
US-08-453-265-5/c
; Sequence 5, Application US/08453265
; Patent No. 5693757
;  GENERAL INFORMATION:
;    APPLICANT:  MacDonald, Marcy E.
;    APPLICANT:  Ambrose, Christine M.
;    APPLICANT:  Duyao, Mabel P.
;    APPLICANT:  Gusella, James F.
;    TITLE OF INVENTION:  Huntingtin DNA, Protein And Uses Thereof
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox
;      STREET:  1100 New York Avenue
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/453,265
;      FILING DATE:  30-MAY-1995
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ludwig, Steven R.
;      REGISTRATION NUMBER:  36,203
;      REFERENCE/DOCKET NUMBER:  0609.3880003
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10366 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  316..9748
US-08-453-265-5


  Query Match             4.4%;  Score 38.2;  DB 2;  Length 10366;
  Best Local Similarity   51.6%;  Pred. No. 0.15;
  Matches  111;  Conservative    0;  Mismatches  103;  Indels    1;  Gaps    1;

Qy      74 GCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTG 133
           ||   ||||| || |      | |  ||||||  |  |||   ||| || | ||   |
Db     552 GCCGGGTGGCGGCGGGGGCGGCGGCGGGGGCGGCTGCGGCTGAGGCAGCAGCGGCTGTGC 493

Qy     134 CGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCGTTTAT-TCAGCAG 192
           | |    || ||| || |||  || | ||||||||  |||  | | ||    |  | || |
Db     492 CTGCGGCGGCGGCTGAGGAAGCTGAGGAGGCGGCGGCGGCGGCGGCGGTGGCGGCTG 433

Qy     193 GCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTGATTGGCAAGGCG 252
             |||    | | |  ||  ||   |||||  || | |  ||  |||| |      |  |
Db     432 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 373

Qy     253 ATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAA 287
           |    |||  || || ||| | |  | |  | |
Db     372 CTGCTGGAAGGACTTGAGGGACTCGAAGGCCTTCA 338


RESULT   5
US-08-457-273B-41/c
; Sequence 41, Application US/08457273B
; Patent No. 5849995
;  GENERAL INFORMATION:
;    APPLICANT:  Hayden, Michael
;    APPLICANT:  Lin, Biaoyang
;    APPLICANT:  Nasir, Jamal
;    TITLE OF INVENTION:  Mouse Model for Huntington's Disease and
;    TITLE OF INVENTION:  Related DNA Sequences
;    NUMBER OF SEQUENCES:  42
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Virginia Bennett
;      STREET:  PO Box 37428
;      CITY:  Raleigh
;      STATE:  No. 5849995th Carolina
;      COUNTRY:  US
;      ZIP:  27627
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/457,273B
;      FILING DATE:
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bennett, Virginia C.
;      REGISTRATION NUMBER:  37,092
;      REFERENCE/DOCKET NUMBER:  3477-85A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-854-1400
;      TELEFAX:  919-854-1401
;  INFORMATION FOR SEQ ID NO:  41:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10348 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-457-273B-41


  Query Match             4.4%;  Score 38.2;  DB 3;  Length 10348;
  Best Local Similarity   51.6%;  Pred. No. 0.15;
  Matches  111;  Conservative    0;  Mismatches  103;  Indels    1;  Gaps    1;

Qy      74 GCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTG 133
           ||   ||||| || |      | |  ||||||  |  |||   ||| || | ||   |
Db     552 GCCGGGTGGCGGCGGGGGCGGCGGCGGGGGCGGCTGCGGCTGAGGCAGCAGCGGCTGTGC 493

Qy     134 CGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCGTTTAT-TCAGCAG 192
           | |    || ||| || |||  || | ||||||||  |||  | | ||    |  | || |
Db     492 CTGCGGCGGCGGCTGAGGAAGCTGAGGAGGCGGCGGCGGCGGCGGCGGTGGCGGCTG 433

Qy     193 GCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTGATTGGCAAGGCG 252
             |||    | | |  ||  ||   |||||  || | |  ||  |||| |      |  |
Db     432 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 373

Qy     253 ATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAA 287
           |    |||  || || ||| | |  | |  | |
Db     372 CTGCTGGAAGGACTTGAGGGACTCGAAGGCCTTCA 338
```

```
Db    432 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 373

Qy    253 ATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAA 287
Db    372 CTGCTGGAAGGACTTGAGGGACTCGAAGGCCTTCA 338


RESULT   6
US-08-433-854-3/c
; Sequence 3, Application US/08433854
; Patent No. 5721119
;  GENERAL INFORMATION:
;    APPLICANT:  Singh, Mohan Bir
;    APPLICANT:  Knox, Robert B.
;    APPLICANT:  Smith, Penelope
;    APPLICANT:  Avjioglu, Asil
;    APPLICANT:  Theerakulpisut, Piyada
;    APPLICANT:  Hough, Terryn
;    APPLICANT:  Suphioglu, Cenk
;    APPLICANT:  Ong, Eng Kok
;    TITLE OF INVENTION:  Ryegrass Pollen Allergen
;    NUMBER OF SEQUENCES:  53
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock Washburn Kurtz Mackiewicz & No. 5721119ris
;      STREET:  One Liberty Place - 46th Floor
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/433,854
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/930,060
;      FILING DATE:  14-AUG-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hohenschutz, Liza D.
;      REGISTRATION NUMBER:  33,712
;      REFERENCE/DOCKET NUMBER:  IMPH-0024
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 568-3100
;      TELEFAX:  (215) 568-3439
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1295 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  25..1041
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  100..1041
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  25..99
US-08-433-854-3


  Query Match             4.1%;  Score 35.8;  DB 2;  Length 1295;
  Best Local Similarity   59.2%;  Pred. No. 0.29;
  Matches   61;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy     95 CCGGTGGGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTGGCCGATGAAA 154

Db   1027 CAGCAGCGGCGGGGGTTGCGGTGGCAGCGGCGGGGGTTGCGGTGGCGGTGGCAGCGGCGG 968

Qy    155 TTGTGAAGGCGGCGACGGGAGTGTCGTTTATTCAGCAGGCGCT 197
Db    967 GGGTAGCGGTGGCGTAGGCAGCGGCGGGGGTTGCGGTGGCGGT 925


RESULT   7
US-08-209-747-1
; Sequence 1, Application US/08209747
; Patent No. 5733771
;  GENERAL INFORMATION:
;    APPLICANT:  Lewis, Randolph V.
;    APPLICANT:  Colgin, Mark
;    TITLE OF INVENTION:  cDNAs Encoding Minor Ampullate Spider
;    TITLE OF INVENTION:  Silk Proteins
;    NUMBER OF SEQUENCES:  56
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Birch, Stewart, Kolasch & Birch
;      STREET:  P.O. Box 747
;      CITY:  Falls Church
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22040-3487
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/209,747
;      FILING DATE:  14-MAR-1994
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Murphy Jr., Gerald M.
;      REGISTRATION NUMBER:  28,977
;      REFERENCE/DOCKET NUMBER:  1447-104P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-205-8000
;      TELEFAX:  703-205-8050
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2793 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Nephila clavipes
;      TISSUE TYPE:  minor ampullate gland
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  183..2675
;      OTHER INFORMATION:  /product= "N. clavipes minor
;      OTHER INFORMATION:  ampullate silk protein"
US-08-209-747-1


  Query Match             4.1%;  Score 35.8;  DB 2;  Length 2793;
  Best Local Similarity   55.1%;  Pred. No. 0.41;
  Matches   70;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy    109 CTGGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCG 168
Db    454 CAGGAGCTGCCGGCGCTGGAGGTTACGGAAGAGGTGCTGGAGGCTACGGTGGACAAGGAG 513

Qy    169 ACGGGAGTGTCGTTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCG 228
Db    514 GATATGGTGCCGGAGCCGGAGCTGGTGCTGCTGCAGCTGCTGGAGCAGGAGCCGGAGGCG 573

Qy    229 TTAATGG 235
```

```
Db    574 CTGGTGG 580


RESULT   8
US-08-174-745A-3/c
; Sequence 3, Application US/08174745A
; Patent No. 5736362
;  GENERAL INFORMATION:
;    APPLICANT:  Singh, Mohan Bir
;    APPLICANT:  Knox, Robert B.
;    APPLICANT:  Smith, Penelope
;    APPLICANT:  Avjioglu, Asil
;    APPLICANT:  Theerakulpisut, Piyada
;    APPLICANT:  Hough, Terryn
;    APPLICANT:  Suphioglu, Cenk
;    APPLICANT:  Ong, Eng Kok
;    TITLE OF INVENTION:  Ryegrass Pollen Allergen
;    NUMBER OF SEQUENCES:  53
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock Washburn Kurtz Mackiewicz & No. 5736362ris
;      STREET:  One Liberty Place - 46th Floor
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/174,745A
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/930,060
;      FILING DATE:  14-AUG-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hohenschutz, Liza D.
;      REGISTRATION NUMBER:  33,712
;      REFERENCE/DOCKET NUMBER:  IMPH-0024
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 568-3100
;      TELEFAX:  (215) 568-3439
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1295 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  25..1041
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  100..1041
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  25..99
US-08-174-745A-3


  Query Match             4.1%;  Score 35.8;  DB 2;  Length 1295;
  Best Local Similarity   59.2%;  Pred. No. 0.29;
  Matches   61;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy      95 CCGGTGGGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTGGCCGATGAAA 154
           | |  | |||| |  | ||| |||| |||| |||  |||||||   |||||| |  |
Db    1027 CAGCAGCGGCGGGGGTTGCGGTGGCAGCGGCGGGGGTTGCGGTGGCGGTGGCAGCGGCGG 968

Qy     155 TTGTGAAGGCGGCGACGGGAGTGTCGTTTATTCAGCAGGCGCT 197
             ||   || ||||  || || | ||    ||  |  |||| |
```

```
Db    967 GGGTAGCGGTGGCGTAGGCAGCGGCGGGGGTTGCGGTGGCGGT 925


RESULT   9
US-08-458-298-1
; Sequence 1, Application US/08458298
; Patent No. 5756677
;  GENERAL INFORMATION:
;    APPLICANT:  Lewis, Randolph V.
;    APPLICANT:  Colgin, Mark
;    TITLE OF INVENTION:  cDNAs Encoding Minor Ampullate Spider
;    TITLE OF INVENTION:  Silk Proteins
;    NUMBER OF SEQUENCES:  56
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Birch, Stewart, Kolasch & Birch
;      STREET:  P.O. Box 747
;      CITY:  Falls Church
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22040-3487
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/458,298
;      FILING DATE:  02-JUN-1995
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/209,747
;      FILING DATE:  14-MAR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Murphy Jr., Gerald M.
;      REGISTRATION NUMBER:  28,977
;      REFERENCE/DOCKET NUMBER:  1447-104P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-205-8000
;      TELEFAX:  703-205-8050
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2793 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Nephila clavipes
;      TISSUE TYPE:  minor ampullate gland
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  183..2675
;      OTHER INFORMATION:  /product= "N. clavipes minor
;      OTHER INFORMATION:  ampullate silk protein"
US-08-458-298-1


  Query Match             4.1%;  Score 35.8;  DB 2;  Length 2793;
  Best Local Similarity   55.1%;  Pred. No. 0.41;
  Matches   70;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy     109 CTGGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCG 168
           | || |||| | ||| |||   || ||| |  ||||   || |  |  ||| |   || |
Db     454 CAGGAGCTGCCGGCGCTGGAGGTTACGGAAGAGGTGCTGGAGGCTACGGTGGACAAGGAG 513

Qy     169 ACGGGAGTGTCGTTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCG 228
               ||| ||       ||| || |||    | | |  |||||| | || | || ||
Db     514 GATATGGTGCCGGAGCCGGAGCTGGTGCTGCTGCAGCTGCTGGAGCAGGAGCCGGAGGCG 573

Qy     229 TTAATGG 235
           |  |||
Db     574 CTGGTGG 580
```

```
RESULT 10
US-08-195-947-3/c
; Sequence 3, Application US/08195947
; Patent No. 5840316
;  GENERAL INFORMATION:
;    APPLICANT:  Singh, Mohan Bir
;    APPLICANT:  Knox, Robert B.
;    APPLICANT:  Smith, Penelope
;    APPLICANT:  Avjioglu, Asil
;    APPLICANT:  Theerakulpisut, Piyada
;    APPLICANT:  Hough, Terryn
;    APPLICANT:  Suphioglu, Cenk
;    APPLICANT:  Ong, Eng Kok
;    TITLE OF INVENTION:  Ryegrass Pollen Allergen
;    NUMBER OF SEQUENCES:  53
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock Washburn Kurtz Mackiewicz & No. 5840316ris
;      STREET:  One Liberty Place - 46th Floor
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/195,947
;      FILING DATE:  14-FEB-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/930,060
;      FILING DATE:  14-AUG-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hohenschutz, Liza D.
;      REGISTRATION NUMBER:  33,712
;      REFERENCE/DOCKET NUMBER:  IMPH-0024
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 568-3100
;      TELEFAX:  (215) 568-3439
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1295 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  25..1041
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  100..1041
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  25..99
US-08-195-947-3


  Query Match             4.1%;  Score 35.8;  DB 3;  Length 1295;
  Best Local Similarity   59.2%;  Pred. No. 0.29;
  Matches   61;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy      95 CCGGTGGGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTGGCCGATGAAA 154
           | |  | |||| |  | ||| |||| |||| |||  |||||||   |||||| |  |
Db    1027 CAGCAGCGGCGGGGGTTGCGGTGGCAGCGGCGGGGGTTGCGGTGGCGGTGGCAGCGGCGG 968

Qy     155 TTGTGAAGGCGGCGACGGGAGTGTCGTTTATTCAGCAGGCGCT 197
             ||   || ||||  || || | ||    ||  |  |||| |
Db     967 GGGTAGCGGTGGCGTAGGCAGCGGCGGGGGTTGCGGTGGCGGT 925


RESULT 11
US-08-433-885-3/c
; Sequence 3, Application US/08433885
; Patent No. 5869333
;  GENERAL INFORMATION:
;    APPLICANT:  Singh, Mohan Bir
;    APPLICANT:  Knox, Robert B.
;    APPLICANT:  Smith, Penelope
;    APPLICANT:  Avjioglu, Asil
;    APPLICANT:  Theerakulpisut, Piyada
;    APPLICANT:  Hough, Terryn
;    APPLICANT:  Suphioglu, Cenk
;    APPLICANT:  Ong, Eng Kok
;    TITLE OF INVENTION:  Ryegrass Pollen Allergen
;    NUMBER OF SEQUENCES:  53
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lahive & Cockfield
;      STREET:  60 State Street, Suite 510
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/433,885
;      FILING DATE:  02-MAY-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mandragouras, Amy E.
;      REGISTRATION NUMBER:  36,207
;      REFERENCE/DOCKET NUMBER:  IMI-039C2D2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 227-7400
;      TELEFAX:  (617) 227-5941
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1295 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  25..1041
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  100..1041
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  25..99
US-08-433-885-3


  Query Match             4.1%;  Score 35.8;  DB 3;  Length 1295;
  Best Local Similarity   59.2%;  Pred. No. 0.29;
  Matches   61;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy      95 CCGGTGGGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTGGCCGATGAAA 154
           | |  | |||| |  | ||| |||| |||| |||  |||||||   |||||| |  |
Db    1027 CAGCAGCGGCGGGGGTTGCGGTGGCAGCGGCGGGGGTTGCGGTGGCGGTGGCAGCGGCGG 968

Qy     155 TTGTGAAGGCGGCGACGGGAGTGTCGTTTATTCAGCAGGCGCT 197
             ||   || ||||  || || | ||    ||  |  |||| |
Db     967 GGGTAGCGGTGGCGTAGGCAGCGGCGGGGGTTGCGGTGGCGGT 925


RESULT 12
```

```
US-08-146-930-1
; Sequence 1, Application US/08146930
; Patent No. 5958764
;  GENERAL INFORMATION:
;    APPLICANT:  Roop, Dennis R.
;    APPLICANT:  Rothnagel, Joseph A.
;    APPLICANT:  Greenhalgh, David A.
;    TITLE OF INVENTION:  SPECIFIC EXPRESSION VECTORS
;    TITLE OF INVENTION:  AND METHODS OF USE
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LYON & LYON
;      STREET:  611 West Sixth Street
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90017
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb storage
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  IBM MS-DOS (Version 5.0)
;      SOFTWARE:  WordPerfect (Version 5.1)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/146,930
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;    PRIOR APPLICATION DATA:  including application
;    PRIOR APPLICATION DATA:  described below:              two
;      APPLICATION NUMBER:  07/876,286
;      FILING DATE:  April 30, 1992
;      APPLICATION NUMBER:  No. 5958764 yet assigned (204/132)
;      FILING DATE:  October 29, 1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warburg, Richard J.
;      REGISTRATION NUMBER:  32,327
;      REFERENCE/DOCKET NUMBER:  204/152
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6530 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-146-930-1

  Query Match             4.0%;  Score 34.8;  DB 4;  Length 6530;
  Best Local Similarity   44.3%;  Pred. No. 1.1;
  Matches  141;  Conservative    0;  Mismatches  177;  Indels    0;  Gaps    0;

Qy      76 GTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTGCG 135
           || | ||| || | | |  | |  | |||| ||  ||  ||| |||| ||    | |
Db    3148 GTGGCGGCGGCGGCTCTAGCTGCGGCGGCGGCTCCTCCGGGGGCGGCGGCGGCGGCTCCA 3207

Qy     136 GTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCGTTTATTCAGCAGGCG 195
           |    || ||| ||| |    | |  |||  | || |  | | |   | |  ||  |||
Db    3208 GCTGCGGAGGCGGATCAGGAGGCGGCGGCTCCTACTGCGGAGGCTCCTCTGGAGGCGGCA 3267

Qy     196 CTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTGATTGGCAAGGCGATT 255
                 |  | || |   || | | || ||  ||      ||| | |    || | |
Db    3268 GCTCCGGTGGCTGCGGCGGCGGTTCCGGAGGCGGCAAGTACTCTGGTGGCGGCGGTGGCT 3327

Qy     256 ACCAAAGCGCTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAGATGGCCGGCAGCATT 315
            |   |||  ||  |   |    |  |    | || | || | | |       |
Db    3328 CCAGCTGCGGAGGCGGCTATTCCGGCGGCGGTGGAAGCAGCGGCGGCTCTAGCTGTGGCG 3387

Qy     316 GTTGGTGCGATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTGGTGGTCGCAGTTGTC 375
           |  | | |    || || |    | |    | |||||     | ||   |  | || |
Db    3388 GCGGCTACTCAGGTGGCGGTGGATCCAGCTGCGGCGGCGGCGGCGGCTATTCCGGTGGCG 3447

Qy     376 GGGAAAGGCGCGGCGGCG 393
           | |  | | || |||| |
Db    3448 GCGGCACGAGCTGCGGAG 3465


RESULT  13
PCT-US93-03993-1
; Sequence 1, Application PC/TUS9303993
;  GENERAL INFORMATION:
;    APPLICANT:  Roop, Dennis R.
;    APPLICANT:  Rothnagel, Joseph A.
;    APPLICANT:  Greenhalgh, David A.
;    TITLE OF INVENTION:  CONSTITUTIVE AND INDUCIBLE EPIDERMAL
;    TITLE OF INVENTION:  VECTOR SYSTEMS
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fulbright & Jaworski
;      STREET:  1301 McKinney, Suite 5100
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  U.S.A.
;      ZIP:  77010-3095
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/03993
;      FILING DATE:  19930428
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Paul, Thomas D.
;      REGISTRATION NUMBER:  32,714
;      REFERENCE/DOCKET NUMBER:  D-5405
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  713/651-5325
;      TELEFAX:  713/651-5246
;      TELEX:  762829
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6530 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
PCT-US93-03993-1

  Query Match             4.0%;  Score 34.8;  DB 5;  Length 6530;
  Best Local Similarity   44.3%;  Pred. No. 1.1;
  Matches  141;  Conservative    0;  Mismatches  177;  Indels    0;  Gaps    0;

Qy      76 GTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTGCG 135
           || | ||| || | | |  | |  | |||| ||  ||  ||| |||| ||    | |
Db    3148 GTGGCGGCGGCGGCTCTAGCTGCGGCGGCGGCTCCTCCGGGGGCGGCGGCGGCGGCTCCA 3207

Qy     136 GTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCGTTTATTCAGCAGGCG 195
           |    || ||| ||| |    | |  |||  | || |  | | |   | |  ||  |||
Db    3208 GCTGCGGAGGCGGATCAGGAGGCGGCGGCTCCTACTGCGGAGGCTCCTCTGGAGGCGGCA 3267

Qy     196 CTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTGATTGGCAAGGCGATT 255
                 |  | || |   || | | || ||  ||      ||| | |    || | |
Db    3268 GCTCCGGTGGCTGCGGCGGCGGTTCCGGAGGCGGCAAGTACTCTGGTGGCGGCGGTGGCT 3327

Qy     256 ACCAAAGCGCTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAGATGGCCGGCAGCATT 315
            |   |||  ||  |   |    |  |    | || | || | | |       |
Db    3328 CCAGCTGCGGAGGCGGCTATTCCGGCGGCGGTGGAAGCAGCGGCGGCTCTAGCTGTGGCG 3387
```

```
Qy    316 GTTGGTGCGATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTGGTGGTCGCAGTTGTC 375
          |  | | |    || || |    | |    | |||||     | ||   |  | || |
Db   3388 GCGGCTACTCAGGTGGCGGTGGATCCAGCTGCGGCGGCGGCGGCGGCTATTCCGGTGGCG 3447

Qy    376 GGGAAAGGCGCGGCGGCG 393
          | |  | | || |||| |
Db   3448 GCGGCACGAGCTGCGGAG 3465


RESULT 14
US-08-147-777-1
; Sequence 1, Application US/08147777
; Patent No. 5914265
;  GENERAL INFORMATION:
;    APPLICANT:  Roop, Dennis R.
;    APPLICANT:  Rothnagel, Joseph A.
;    APPLICANT:  Greenhalgh, David A.
;    APPLICANT:  Yuspa, Stuart H.
;    TITLE OF INVENTION:  KERATIN K1 EXPRESSION VECTORS
;    TITLE OF INVENTION:  AND METHODS OF USE
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LYON & LYON
;      STREET:  611 West Sixth Street
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90017
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb storage
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  IBM MS-DOS (Version 5.0)
;      SOFTWARE:  WordPerfect (Version 5.1)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/147,777
;      FILING DATE:
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;    PRIOR APPLICATION DATA:  including application
;    PRIOR APPLICATION DATA:  described below:           two
;      APPLICATION NUMBER:  07/876,289
;      FILING DATE:  April 30, 1992
;      APPLICATION NUMBER:  Unassigned (204/144)
;      FILING DATE:  October 29, 1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warburg, Richard J.
;      REGISTRATION NUMBER:  32,327
;      REFERENCE/DOCKET NUMBER:  204/153
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10747 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-147-777-1


  Query Match             3.9%;  Score 34.2;  DB 4;  Length 10747;
  Best Local Similarity   46.4%;  Pred. No. 2;
  Matches  111;  Conservative    0;  Mismatches  128;  Indels    0;  Gaps    0;

Qy     50 TCGGCGCGCTGCTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTC 109
          |||| |   || ||||| | ||    ||    ||  ||  |     |   | |
Db   1497 TCGGAGTCTTGTTAACCTTGGTGGCAGTAAAAGCATCTCCATAAGTGTGGCTAGAGGAGG 1556

Qy    110 TGGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGA 169
          |||   |||  | ||      | | ||| |||||||  |  |   | |||  || |||
Db   1557 TGGACGTGGTAGTGGCTTTGGTGGTGGTTATGGTGGTGGTGGCTTTGGTGGTGGTGGCTT 1616
```

```
Qy    170 CGGGAGTGTCGTTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGT 229
          ||  |||  |  | |   | ||| |  |     | |  |||     ||| |   | |
Db   1617 TGGTGGTGGTGGCTTTGGTGGAGGTGGCATTGGGGGTGGTGGCTTTGGTGGTTTTGGCAG 1676

Qy    230 TAATGGAGCTGATTGGCAAGGCGATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAAG 288
          |  ||| | || ||     || | |  |    | | |||  |   | || || | || |
Db   1677 TGGTGGTGGTGGTTTTGGTGGAGGTGGCTTTGGGGGTGGTGGATATGGGGGTGGTTATG 1735


RESULT 15
US-08-147-777-3
; Sequence 3, Application US/08147777
; Patent No. 5914265
;  GENERAL INFORMATION:
;    APPLICANT:  Roop, Dennis R.
;    APPLICANT:  Rothnagel, Joseph A.
;    APPLICANT:  Greenhalgh, David A.
;    APPLICANT:  Yuspa, Stuart H.
;    TITLE OF INVENTION:  KERATIN K1 EXPRESSION VECTORS
;    TITLE OF INVENTION:  AND METHODS OF USE
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LYON & LYON
;      STREET:  611 West Sixth Street
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90017
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb storage
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  IBM MS-DOS (Version 5.0)
;      SOFTWARE:  WordPerfect (Version 5.1)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/147,777
;      FILING DATE:
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;    PRIOR APPLICATION DATA:  including application
;    PRIOR APPLICATION DATA:  described below:           two
;      APPLICATION NUMBER:  07/876,289
;      FILING DATE:  April 30, 1992
;      APPLICATION NUMBER:  Unassigned (204/144)
;      FILING DATE:  October 29, 1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warburg, Richard J.
;      REGISTRATION NUMBER:  32,327
;      REFERENCE/DOCKET NUMBER:  204/153
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  24979 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-147-777-3


  Query Match             3.9%;  Score 34.2;  DB 4;  Length 24979;
  Best Local Similarity   46.4%;  Pred. No. 2.9;
  Matches  111;  Conservative    0;  Mismatches  128;  Indels    0;  Gaps    0;

Qy     50 TCGGCGCGCTGCTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTC 109
          |||| |   || ||||| | ||    ||    ||  ||  |     |   | |
Db  15729 TCGGAGTCTTGTTAACCTTGGTGGCAGTAAAAGCATCTCCATAAGTGTGGCTAGAGGAGG 15788

Qy    110 TGGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGA 169
          |||   |||  | ||      | | ||| |||||||  |  |   | |||  || |||
```

```
Db  15789 TGGACGTGGTAGTGGCTTTGGTGGTGGTTATGGTGGTGGTGGCTTTGGTGGTGGTGGCTT 15848

Qy    170 CGGGAGTGTCGTTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGT 229
          ||  |||  |  | |  | ||| |  |     | |  |||     ||| |   | |
Db  15849 TGGTGGTGGTGGCTTTGGTGGAGGTGGCATTGGGGGTGGTGGCTTTGGTGGTTTTGGCAG 15908

Qy    230 TAATGGAGCTGATTGGCAAGGCGATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAAG 288
          |  ||| | || ||     || | |  |    | | |||  |   | || || | || |
Db  15909 TGGTGGTGGTGGTTTTGGTGGAGGTGGCTTTGGGGGTGGTGGATATGGGGGTGGTTATG 15967


Search completed: October  5, 1999, 18:00:38
Job time: 11478 sec
```

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 1999, 14:49:18 ; Search time 223.92 Seconds
(without alignments)
972.075 Million cell updates/sec

Title: US-09-068-804-1
Perfect score: 870
Sequence: 1 CGCAAAGCCGAGGAAACGAA.....TGCGCCAGAGTCGCGCATAA 870

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	870	100.0	3441	1	T67035	Salmonella virulen
	2	868.4	99.8	870	1	T67028	Salmonella secrete
	3	178.6	20.5	1743	1	T85939	Shigella flexneri
c	4	49.4	5.7	1648	1	Q25532	Sequence of genomi
	5	45.2	5.2	1505	1	Q55750	Genomic clone G11F
	6	43.6	5.0	4018	1	Q63879	Polyhydroxyalkanoa
c	7	38.2	4.4	10366	1	Q67401	Huntingtin DNA\IT1
c	8	38.2	4.4	10366	1	T97924	Composite sequence
c	9	38.2	4.4	10366	1	V05828	Human huntingtin c
c	10	38.2	4.4	515	1	X18930	Human huntingtin g
c	11	38	4.4	342	1	Q31795	PRP2 gene. Isolate
c	12	37.8	4.3	5163	1	V20700	Cryptosporidium pa
c	13	37.4	4.3	1107	1	V38816	DNA encoding a dur
	14	37	4.3	864	1	Q03877	Merozoite surface
c	15	36.2	4.2	5318	1	V20701	Cryptosporidium pa
c	16	35.8	4.1	1295	1	Q38230	Sequence of rye gr
	17	35.8	4.1	2744	1	Q98470	MiSP1-containing p
	18	35.6	4.1	1491	1	X15147	Nucleotide sequenc
	19	35	4.0	2440	1	Q06469	Sequence encoding
	20	35	4.0	3086	1	Q06470	Sequence encoding
c	21	34.8	4.0	836	1	Q12218	SMUC-41 intestinal
	22	34.8	4.0	6530	1	Q51557	Loricrin gene. Con
	23	34.2	3.9	202	1	N81094	Sequence of a synt
	24	34	3.9	2338	1	Q14183	N.clavipes draglin
	25	34	3.9	2338	1	V23249	Nephila clavipes s
	26	33.8	3.9	334	1	V88505	EST clone EN108. N
	27	33.8	3.9	114955	1	X53491	Human adenosine A1
c	28	33.6	3.9	910	1	X30944	Streptococcus pneu
c	29	33.6	3.9	730	1	X39791	Gastric cancer ass
c	30	33.4	3.8	860	1	Q06486	Intestinal mucin c
c	31	33.4	3.8	859	1	Q12220	SMUC-53 intestinal
	32	33.4	3.8	3319	1	Q70152	K39 gene of Leishm
	33	33.4	3.8	3319	1	T42166	Leishmania chagasi
c	34	33.2	3.8	2291	1	N80309	Entire amylase gen
c	35	33.2	3.8	2271	1	Q06844	Amylase gene from
	36	33.2	3.8	1320	1	X24723	Pineapple polyphen
	37	33	3.8	15464	1	Q11415	Ryanodine receptor
	38	33	3.8	15672	1	Q10613	Rianodin receptor
c	39	32.6	3.7	1064	1	Q31797	PRP3 gene (genomic
c	40	32.4	3.7	836	1	Q06484	Intestinal mucin c
	41	32.4	3.7	1951	1	T30566	Eukaryotic cell gr
	42	32.4	3.7	564	1	V32661	Human Thioredoxin
	43	32.2	3.7	267	1	Q60366	Human brain Expres
c	44	32.2	3.7	9757	1	Q97686	Infectious rubella
c	45	32.2	3.7	9759	1	T89642	Infectious rubella

ALIGNMENTS

RESULT 1
T67035
ID T67035 standard; DNA; 3441 BP.
AC T67035;
DT 04-AUG-1997 (first entry)
DE Salmonella virulence factor SspBCDA DNA region.
KW Salmonella secreted protein; SspB; SspC; SspD; SspA;
KW bacterial-mediated endocytosis; diagnosis; therapy; vaccine;
KW attenuation; virulence; ss.
OS Salmonella typhimurium.
FH Key Location/Qualifiers
FT cds 1. .870
FT /*tag= a
FT /product= SspB (incomplete)
FT cds 898. .2127
FT /*tag= b
FT /product= SspC
FT cds 2198. .3228
FT /*tag= c
FT /product= SspD
FT cds 3247. .3441
FT /*tag= d
FT /product= SspA (incomplete)
PN WO9718225-A1.
PD 22-MAY-1997.
PF 14-NOV-1996; U18504.
PR 14-NOV-1995; US-006733.
PA (GEHO) GEN HOSPITAL CORP.
PI Miller SI;
DR WPI; 97-289217/26.
DR P-PSDB; W15268-71.
PT New isolated Salmonella secreted proteins and related genes - used
PT to develop products for the detection, treatment or prevention of
PT Salmonella infections
PS Claim 11; Fig 28; 95pp; English.
CC A DNA molecule (T67035) comprises the SspBCDA locus (truncated at
CC the 5' and 3' ends) including the partial sspB, sspC, sspD and
CC partial sspA sequences (see also T67028-31) coding for Salmonella
CC typhimurium secreted proteins SspB, SspC, SspD and SspA (W15268-
CC 70). To identify the ssp genes, S. typhimurium mutants were
CC investigated in invasion assays on HEp-2 epithelial cells. 2
CC Ssp mutants with transposon insertions located between spaT and
CC prgH (see also T67036) were identified. One mutant lacked Ssp of
CC 87, 42 and 36 kDa. The invasiveness of this mutant implicated the
CC 42 and 36 kDa Ssp in invasion. DNA sequencing of this region
CC identified the 4 ssp (partial) genes. The ssp genes can be used in
CC methods for diagnosing Salmonella infection and in methods of
CC targetting an antigen to epithelial cells and of inducing a
CC cytotoxic T cell immune response in a mammal. Ssp gene mutation
CC can be used to create attenuated Salmonella strains useful as live
CC vaccines.
SQ Sequence 3441 BP; 978 A; 761 C; 925 G; 777 T;

Query Match 100.0%; Score 870; DB 1; Length 3441;
Best Local Similarity 100.0%; Pred. No. 1.3e-240;
Matches 870; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 CGCAAAGCCGAGGAAACGAACCGCATTATGGGATGTATCGGGAAAGTCCTCGGCGCGCTG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CGCAAAGCCGAGGAAACGAACCGCATTATGGGATGTATCGGGAAAGTCCTCGGCGCGCTG 60

Qy      61 CTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCT 120
```

```
Qy    121 GCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCG 180

Qy    181 TTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTG 240

Qy    241 ATTGGCAAGGCGATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ATTGGCAAGGCGATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAG 300

Qy    301 ATGGCCGGCAGCATTGTTGGTGCGATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 ATGGCCGGCAGCATTGTTGGTGCGATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTG 360

Qy    361 GTGGTCGCAGTTGTCGGGAAAGGCGCGGCGGCGAAACTGGGTAACGCGCTGAGCAAAATG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GTGGTCGCAGTTGTCGGGAAAGGCGCGGCGGCGAAACTGGGTAACGCGCTGAGCAAAATG 420

Qy    421 ATGGGCGAAACGATTAAGAAGTTGGTGCCTAACGTGCTGAAACAGTTGGCGCAAAACGGC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 ATGGGCGAAACGATTAAGAAGTTGGTGCCTAACGTGCTGAAACAGTTGGCGCAAAACGGC 480

Qy    481 AGCAAACTCTTTACCCAGGGGATGCAACGTATTACTAGCGGTCTGGGTAATGTGGGTAGC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 AGCAAACTCTTTACCCAGGGGATGCAACGTATTACTAGCGGTCTGGGTAATGTGGGTAGC 540

Qy    541 AAGATGGGCCTGCAAACGAATGCCTTAAGTAAAGAGCTGGTAGGTAATACCCTAAATAAA 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 AAGATGGGCCTGCAAACGAATGCCTTAAGTAAAGAGCTGGTAGGTAATACCCTAAATAAA 600

Qy    601 GTGGCGTTGGGCATGGAAGTCACGAATACCGCAGCCCAGTCAGCCGGTGGTGTTGCCGAG 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GTGGCGTTGGGCATGGAAGTCACGAATACCGCAGCCCAGTCAGCCGGTGGTGTTGCCGAG 660

Qy    661 GGCGTATTTATTAAAAATGCCAGCGAGGCGCTTGCTGATTTTATGCTCGCCCGTTTTGCC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 GGCGTATTTATTAAAAATGCCAGCGAGGCGCTTGCTGATTTTATGCTCGCCCGTTTTGCC 720

Qy    721 ATGGATCAGATTCAGCAGTGGCTTAAACAATCCGTAGAAATATTTGGTGAAAACCAGAAG 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 ATGGATCAGATTCAGCAGTGGCTTAAACAATCCGTAGAAATATTTGGTGAAAACCAGAAG 780

Qy    781 GTAACGGCGGAACTGCAAAAAGCCATGTCTTCTGCGGTACAGCAAAATGCGGATGCTTCG 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GTAACGGCGGAACTGCAAAAAGCCATGTCTTCTGCGGTACAGCAAAATGCGGATGCTTCG 840

Qy    841 CGTTTTATTCTGCGCCAGAGTCGCGCATAA 870
          |||||||||||||||||||||||||||||
Db    841 CGTTTTATTCTGCGCCAGAGTCGCGCATAA 870
```

```
RESULT   2
T67028
ID   T67028 standard; DNA; 870 BP.
AC   T67028;
DT   04-AUG-1997  (first entry)
DE   Salmonella secreted protein SspB DNA.
KW   Salmonella secreted protein; SspB; bacterial-mediated endocytosis;
KW   diagnosis; therapy; vaccine; attenuation; virulence; ss.
OS   Salmonella typhimurium.
FH   Key             Location/Qualifiers
FT   cds             1. .870
FT                   /*tag=  a
FT                   /product= SspB (incomplete)
FT                   /note= "deduced amino acid residues from bases
FT                   286-354 of the coding sequence to not
FT                   correspond to the SspB sequence given in
FT                   Fig 22"
PN   WO9718225-A1.
PD   22-MAY-1997.
PF   14-NOV-1996; U18504.
PR   14-NOV-1995; US-006733.
PA   (GEHO ) GEN HOSPITAL CORP.
PI   Miller SI;
DR   WPI; 97-289217/26.
DR   P-PSDB; W15268.
PT   New isolated Salmonella secreted proteins and related genes - used
PT   to develop products for the detection, treatment or prevention of
PT   Salmonella infections
PS   Claim 3; Fig 19; 95pp; English.
CC   A DNA molecule (T67028) codes for (incomplete) Salmonella secreted
CC   protein SspB (W15268), a virulence factor involved in bacterial
CC   mediated endocytosis, which is able to translocate into human
CC   epithelial cells.  To identify Ssp genes, S. typhimurium mutants
CC   were investigated in invasion assays on HEp-2 epithelial cells.  2
CC   Ssp mutants with transposon insertions located between spaT and
CC   prgH were identified.  One mutant lacked Ssp of 87, 42 and 36 kDa.
CC   The invasiveness of this mutant implicated the 42 and 36 kDa Ssp in
CC   invasion.  DNA sequencing of this region (see also T67035)
CC   identified 2 complete and 2 partial open reading frames, designated
CC   sspB, sspC (T67029), sspD (T67030) and sspA (T67031).  These
CC   nucleic acids can be used in methods for diagnosing Salmonella
CC   infection and in methods of targetting an antigen to epithelial
CC   cells and of inducing a cytotoxic T cell immune response in a
CC   mammal.  Ssp gene mutation can be used to create attenuated
CC   Salmonella strains useful as live vaccines.
SQ   Sequence   870 BP;    216 A;    178 C;    281 G;    195 T;
```

Query Match 99.8%; Score 868.4; DB 1; Length 870;
Best Local Similarity 99.9%; Pred. No. 2.1e-240;
Matches 869; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      1 CGCAAAGCCGAGGAAACGAACCGCATTATGGGATGTATCGGGAAAGTCCTCGGCGCGCTG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CGCAAAGCCGAGGAAACGAACCGCATTATGGGATGTATCGGGAAAGTCCTCGGCGCGCTG 60

Qy     61 CTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCT 120

Qy    121 GCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCG 180

Qy    181 TTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTG 240

Qy    241 ATTGGCAAGGCGATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ATTGGCAAGGCGATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAG 300

Qy    301 ATGGCCGGCAGCATTGTTGGTGCGATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 ATGGCCGGCAGCATTGTTGGTGCGATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTG 360

Qy    361 GTGGTCGCAGTTGTCGGGAAAGGCGCGGCGGCGAAACTGGGTAACGCGCTGAGCAAAATG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GTGGTCGCAGTTGTCGGGAAAGGCGCGGCGGCGAAACTGGGTAACGCGCTGAGCAAAATG 420

Qy    421 ATGGGCGAAACGATTAAGAAGTTGGTGCCTAACGTGCTGAAACAGTTGGCGCAAAACGGC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 ATGGGCGAAACGATTAAGAAGTTGGTGCCTAACGTGCTGAAACAGTTGGCGCAAAACGGC 480

Qy    481 AGCAAACTCTTTACCCAGGGGATGCAACGTATTACTAGCGGTCTGGGTAATGTGGGTAGC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 AGCAAACTCTTTACCCAGGGGATGCAACGTATTACTAGCGGTCTGGGTAATGTGGGTAGC 540

Qy    541 AAGATGGGCCTGCAAACGAATGCCTTAAGTAAAGAGCTGGTAGGTAATACCCTAAATAAA 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 AAGATGGGCCTGCAAACGAATGCCTTAAGTAAAGAGCTGGTAGGTAATACCCTAAATAAA 600

Qy    601 GTGGCGTTGGGCATGGAAGTCACGAATACCGCAGCCCAGTCAGCCGGTGGTGTTGCCGAG 660
```

```
Db    601 GTGGCGTTGGGCATGGAAGTCACGAATACCGCAGCCCAGTCAGCCGGTGGTGTTGCCGAG 660

Qy    661 GGCGTATTTATTAAAAATGCCAGCGAGGCGCTTGCTGATTTTATGCTCGCCCGTTTTGCC 720

Db    661 GGCGTATTTATTAAAAATGCCAGCGAGGCGCTTGCTGATTTTATGCTCGCCCGTTTTGCC 720

Qy    721 ATGGATCAGATTCAGCAGTGGCTTAAACAATCCGTAGAAATATTTGGTGAAAACCAGAAG 780

Db    721 ATGGATCAGATTCAGCAGTGGCTTAAACAATCCGTAGAAATATTTGGTGAAAACCAGAAG 780

Qy    781 GTAACGGCGGAACTGCAAAAAGCCATGTCTTCTGCGGTACAGCAAAATGCGGATGCTTCG 840

Db    781 GTAACGGCGGAACTGCAAAAAGCCATGTCTTCTGCGGTACAGCAAAATGCGGATGCTTCG 840

Qy    841 CGTTTTATTCTGCGCCAGAGTCGCGCATAA 870

Db    841 GGTTTTATTCTGCGCCAGAGTCGCGCATAA 870
```

```
RESULT   3
T85939
ID   T85939 standard; DNA; 1743 BP.
AC   T85939;
DT   10-MAR-1998  (first entry)
DE   Shigella flexneri IpaB gene.
KW   IpaB; apoptosis; interleukin-1-beta converting enzyme; ICE; cacner;
KW   autoimmune disease; inflammation; AIDS; degenerative disease;
KW   Alzheimer's disease; myelodysplastic disorder; ischaemic injury;
KW   toxin-induced liver disease; gene therapy; shigellosis; ss.
OS   Shigella flexneri.
PN   WO9726790-A1.
PD   31-JUL-1997.
PF   10-JAN-1997; U00399.
PR   25-JAN-1996; US-591079.
PA   (UYNY ) UNIV NEW YORK STATE.
PI   Chen Y, Zychlinsky A;
DR   WPI; 97-393276/36.
DR   P-PSDB; W29893.
PT   Shigella IpaB protein induces apoptosis by binding to
PT   interleukin-1-beta-converting enzyme - useful for treating cancer,
PT   auto-immunity, inflammation, etc. by eradication of unwanted cells
PS   Claim 3; Page 67-69; 108pp; English.
CC   This sequence comprises DNA encoding the IpaB protein (see W29893)
CC   of Shigella flexneri.  This protein induces apoptosis by binding to
CC   interleukin-1-beta converting enzyme (ICE).  A claimed method of
CC   inducing apoptosis in a eukaryotic cell, thereby killing the cell,
CC   comprises: (a) providing to a cell to be killed a DNA molecule (I)
CC   in expressible form which encodes the Shigella IpaB protein or a
CC   functional derivative; and (b) causing (I) to be expressed in the
CC   cell, thereby producing the IpaB protein or derivative which
CC   induces apoptosis and kills the cell.  This approach is useful in
CC   treating e.g. cancer, autoimmunity, inflammation and chronic viral
CC   infections.  Protein or peptide fragments, and the DNA encoding
CC   them, can also be used to treat diseases that involve apoptotic
CC   mechanisms in their pathogenesis, e.g. AIDS, degenerative diseases
CC   such as Alzheimer's disease, myelodysplastic disorders, ischaemic
CC   injury or toxin-induced liver disease.  Methods of gene therapy
CC   relying upon controlled expression of IpaB in a target cell are
CC   disclosed.  Also claimed are methods for: inhibiting the
CC   interaction of an apoptosis-inducing protein or peptide with ICE;
CC   detecting a compound capable of inhibiting the binding of IpaB to
CC   ICE; screening a candidate protein for its ability to interact
CC   with IpaB in a cell; and treatment of shigellosis.  Homologues of
CC   IpaB such as yopB of Yersinia entercolitica and Yersinia
CC   pseudotuberculosis, and sipB of Salmonella typhimurium and
CC   Salmonella typhi can also be used in the methods of the invention.
SQ   Sequence   1743 BP;    632 A;    360 C;    302 G;    449 T;
```

```
  Query Match            20.5%;  Score 178.6;  DB 1;  Length 1743;
  Best Local Similarity  58.3%;  Pred. No. 6.3e-42;
  Matches  313;  Conservative    0;  Mismatches  224;  Indels    0;  Gaps    0;
```

```
Qy      1 CGCAAAGCCGAGGAAACGAACCGCATTATGGGATGTATCGGGAAAGTCCTCGGCGCGCTG 60

Db    892 CGTAAAGCAGAAGAACTCAACAGAGTAATGGGTTGTGTTGGGAAAATACTTGGGGCACTT 951

Qy     61 CTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCT 120

Db    952 TTAACTATCGTTAGTGTTGTTGCAGCAGCTTTTTCTGGAGGAGCCTCTCTAGCACTGGCA 1011

Qy    121 GCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCG 180

Db   1012 GCTGTTGGTTTAGCTCTTATGGTTACGGATGCTATAGTACAAGCAGCGACCGGCAATTCC 1071

Qy    181 TTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTG 240

Db   1072 TTCATGGAACAAGCCCTGAATCCGATCATGAAAGCAGTCATTGAACCCTTAATCAAACTC 1131

Qy    241 ATTGGCAAGGCGATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAG 300

Db   1132 CTTTCAGATGCATTTACAAAAATGCTCGAAGGCTTGGGCGTCGACTCGAAAAAAGCCAAA 1191

Qy    301 ATGGCCGGCAGCATTGTTGGTGCGATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTG 360

Db   1192 ATGATTGGCTCTATTCTGGGGGCAATCGCAGGCGCTCTTGTCCTAGTTGCAGCAGTCGTT 1251

Qy    361 GTGGTCGCAGTTGTCGGGAAAGGCGCGGCGGCGAAACTGGGTAACGCGCTGAGCAAAATG 420

Db   1252 CTCGTAGCCACTGTTGGTAAACAGGCAGCAGCAAAACTTGCAGAAAATATTGGCAAAATA 1311

Qy    421 ATGGGCGAAACGATTAAGAAGTTGGTGCCTAACGTGCTGAAACAGTTGGCGCAAAACGGC 480

Db   1312 ATAGGTAAAACCCTCACAGACCTTATACCAAAGTTTCTCAAGAATTTTTCTTCTCAACTG 1371

Qy    481 AGCAAACTCTTTACCCAGGGGATGCAACGTATTACTAGCGGTCTGGGTAATGTGGGT 537

Db   1372 GACGATTTAATCACTAATGCTGTTGCCAGATTAAATAAATTTCTTGGTGCAGCGGGT 1428
```

```
RESULT   4
Q25532/c
ID   Q25532 standard; DNA; 1648 BP.
AC   Q25532;
DT   22-NOV-1992  (first entry)
DE   Sequence of genomic clone contg. the entire Histidine-rich
DE   protein (HisRP) gene.
KW   Malaria vaccine; Histidine-rich protein; cytoadherence; ss.
OS   Plasmodium lophurae.
FH   Key             Location/Qualifiers
FT   signal_peptide  219. .359
FT                   /*tag=  a
FT   cds             491. .562
FT                   /*tag=  b
FT                   /product= propeptide
FT   intron          360. .490
FT                   /*tag=  c
FT   misc_feature    60. .64
FT                   /*tag=  d
FT                   /label= splice sequence
FT   misc_feature    485. .490
FT                   /*tag=  e
FT                   /label= splice sequence
FT   mat_peptide     563. .1477
FT                   /*tag=  f
PN   US5116965-A.
PD   25-MAY-1992.
PF   26-AUG-1986; 900401.
PR   26-AUG-1986; US-900401.
PA   (SLOK ) SLOAN KETTERING INST CANCER.
PI   Pologe L, Ravetch JV;
DR   WPI; 92-199590/24.
DR   P-PSDB; R24393.
PT   Histidine-rich protein associated with Plasmodium knob phenotype -
PT   and DNA encoding it, used for in vitro diagnosis of P.
```

PT Falciparum infection.
PS Disclosure; Fig 7A-B; 29pp; English.
CC Two variants of HisRP are produced by P. falciparum. One is
CC associated with what is referred to as "knobby phenotype" (K30) and
CC "knobless phenotype" (K-). The "knobby" and "knobless" phenotypes
CC have been implicated in cytoadherence, which is characteristic of
CC erythrocyte infection. It has now been found that cDNA expressing
CC both K+ and K- HisRP can be obtained by the use of P. lophurae HisRP
CC expressing DNA. The genomic clone (Q25532) is encoded in two
CC exons, separating the signal peptide-encoding sequence from the
CC pro-sequence, confirming that synthesis of the protein occurs via
CC the preproprotein. Oligo. probes synthesised to the signal
CC peptide-encoding exon reveal multiple homologous DNA sequences in
CC the P. lophurae genome. The sequence of mature proteins is arranged
CC in numerous tandem repeats with up to nine histidine residues in a
CC row, similar to other Plasmodium proteins for which sequence data
CC have so far been reported.
SQ Sequence 1648 BP; 569 A; 488 C; 150 G; 441 T;

Query Match 5.7%; Score 49.4; DB 1; Length 1648;
Best Local Similarity 44.9%; Pred. No. 8.6e-05;
Matches 188; Conservative 0; Mismatches 231; Indels 0; Gaps 0;

```
Qy     29 TGGGATGTATCGGGAAAGTCCTCGGCGCGCTGCTAACCATTGTCAGCGTTGTGGCCGCTG 88
Db   1425 TGGTGTGCATCGTGATGATGATGGTGGTGGTGGTGTGCATCGTGATGATGATGGTGGTGG 1366

Qy     89 TTTTTACCGGTGGGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTGGCCG 148
Db   1365 TGATGTGCATCGTGGTGATGGTGGTGGTGGTGTGCATCGTGATGATGATGGTGGTGGTGG 1306

Qy    149 ATGAAATTGTGAAGGCGGCGACGGGAGTGTCGTTTATTCAGCAGGCGCTAAACCCGATTA 208
Db   1305 TGTGCATCGTGATGATGATGGTGGTGGTGGTGTGCATCGTGATGATGATGGTGGTGGTGA 1246

Qy    209 TGGAGCATGTGCTGAAGCCGTTAATGGAGCTGATTGGCAAGGCGATTACCAAAGCGCTGG 268
Db   1245 TGTGCATCGTGATCATGATGGTGGTGGTGGTGTGCATCGTGATGATGATGGTGGTGGTGG 1186

Qy    269 AAGGATTAGGCGTCGATAAGAAAACGGCAGAGATGGCCGGCAGCATTGTTGGTGCGATTG 328
Db   1185 TGTGCATCATGGTGGTGGTGGTGATGATGGTGGTGACCATGGTGGTGGTGGTGATGATGG 1126

Qy    329 TCGCCGCTATTGCCATGGTGGCGGTCATTGTGGTGGTCGCAGTTGTCGGGAAAGGCGCGG 388
Db   1125 TGGTGACCATGGTGGTGGTGGTGATGATGGTGGTGACCATGGTGGTGGTGGTGATGATGG 1066

Qy    389 CGGCGAAACTGGGTAACGCGCTGAGCAAAATGATGGGCGAAACGATTAAGAAGTTGGTG 447
Db   1065 TGTGGAGCATGGTGGTGGTGGTGATGATGGTGTGGAGCATGGTGGTGGTGGTGATGATG 1007
```

RESULT 5
Q55750
ID Q55750 standard; DNA; 1505 BP.
AC Q55750;
DT 23-JUL-1994 (first entry)
DE Genomic clone G11F, includes a Brassica root specific promoter.
KW Plants; transformation; food crops; toxins; insects; fungal;
KW infection; ss.
OS Brassica napus.

FH	Key	Location/Qualifiers
FT	misc_feature	552. .554
FT		/*tag= a
FT		/note= "translation start site"
FT	caat_signal	438. .440
FT		/*tag= b
FT	tata_signal	485. .488
FT		/*tag= c

PN WO9402619-A.
PD 03-FEB-1994.
PF 12-JUL-1993; U06541.
PR 16-JUL-1992; US-915246.
PA (PION-) PIONEER HI-BRED INT INC.
PI Baszczynski CJ, Bellemare G, Boivin R, Fallis L;
DR WPI; 94-048877/06.
PT Brassica root-specific promoter sequence and vectors containing
PT it - for expression of e.g. toxins in the roots of crops
PT susceptible to root diseases
PS Disclosure; Fig 4; 54pp; English.
CC The sequence shows the G11F genomic clone isolated from DNA from
CC the roots of Brassica napus and containing a promoter sequence
CC upstream from the translation start site. The promoter enhances
CC gene expression in roots and can be used to confer disease
CC resistance or immunity upon a susceptible plant root by transforming
CC the plant with a composite gene capable of abundant expression in the
CC gene, comprising the promoter and the gene of interest. This method
CC can be used to transform plants, partic food crops that have roots
CC susceptible to fungal or insect diseases. The roots of the transformed
CC plants express a toxin or substance that provides immunity or resistance
CC to the fungal or insect diseases.
CC See also Q55749 and Q56913-5.
SQ Sequence 1505 BP; 395 A; 201 C; 528 G; 381 T;

Query Match 5.2%; Score 45.2; DB 1; Length 1505;
Best Local Similarity 47.2%; Pred. No. 0.0013;
Matches 137; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

```
Qy    111 GGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGAC 170
Db   1047 GGTGGTGGAGGCGGTGGAAATGGAGGCGGTGGAGGAGGTGGAGGTGCACACGGTGGTGGA 1106

Qy    171 GGGAGTGTCGTTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTT 230
Db   1107 TACGGTGGTGGAGAAGGCGCTGGTGCTGGAGGAGGATATGGAGGTGGCGGTGCAGGTGGA 1166

Qy    231 AATGGAGCTGATTGGCAAGGCGATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAAGAA 290
Db   1167 CATGGAGGTGGTGGAGGCGGGGGAAAAGGAGGCGGTGGAGGAGGAGGTTCTGGCGCCGGT 1226

Qy    291 AACGGCAGAGATGGCCGGCAGCATTGTTGGTGCGATTGTCGCCGCTATTGCCATGGTGGC 350
Db   1227 GGAGCTCACGGTGGTGGTTATGGTGCCGGAGGTGGAGCTGGAGAGGGATACGGTGGTGGT 1286

Qy    351 GGTCATTGTGGTGGTCGCAGTTGTCGGGAAAGGCGCGGCGGCGAAACTGG 400
Db   1287 GGTGGAGAAGGAGGACACGGTGGTGGAGGAGGCGGTGGTGGTGGAGCTGG 1336
```

RESULT 6
Q63879
ID Q63879 standard; DNA; 4018 BP.
AC Q63879;
DT 07-DEC-1994 (first entry)
DE Polyhydroxyalkanoate decomposing activity polypeptide.
KW Polyhydroxyalkanoate decomposing enzyme; PHA; ds.
OS Zoogloea ramigera.

FH	Key	Location/Qualifiers
FT	cds	395. .1153
FT		/*tag= a
FT		/product=
FT		/note= "ORF 1"
FT	cds	3173. .3958
FT		/*tag= b
FT		/product=
FT		/note= "ORF 2"

PN J06086681-A.
PD 29-MAR-1994.
PF 07-SEP-1992; 279099.
PR 07-SEP-1992; JP-279099.
PA (ELED) DENKI KAGAKU KOGYO KK.
PA (UYKA-) GH KANAGAWA DAIGAKU.
PA (CHIK-) ZH CHIKYU KANKYO SANGYO GIJITSU KENKYU.
DR WPI; 94-140539/17.

```
DR   P-PSDB; R53915 AND R63805.
PT   DNA encoding poly-hydroxy-alkanoate decomposing enzyme - useful
PT   for preparing poly-hydroxy-alkanoate in large quantities
PS   Claim 3; Page 9-14; 15pp; Japanese.
CC   Q63879 shows two ORFs, the products of which show
CC   polyhydroxyalkanoate decomposing activity. The method used allows
CC   the polypeptide to be prepared in large quantities.
SQ   Sequence   4018 BP;    767 A;    1298 C;    1305 G;    648 T;


  Query Match             5.0%;  Score 43.6;  DB 1;  Length 4018;
  Best Local Similarity   44.6%;  Pred. No. 0.0059;
  Matches  172;  Conservative    0;  Mismatches  214;  Indels    0;  Gaps    0;

Qy      64 ACCATTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCTGCG 123
           || ||   |  |   || || |    |   |  | |||   | |   ||   |||| ||
Db     392 ACGATGACCCTCCCGGTCGCGGTGACTACGATGGCTGGCATGGGCACGGTCATGGCCACG 451

Qy     124 GTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCGTTT 183
           ||         |||  |||| ||||   |  |  |||  |||||  ||||    | ||
Db     452 GTCACGGGCACGGCCATGGCGGCCACGGCCACGGTGGTGGCGGTCACGGCCACGGCGGCG 511

Qy     184 ATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTGATT 243
             | || |  |  | || | || |   ||    | | || |  |||     ||
Db     512 GCGGTCATGACGGTGGAGGCCATGACGGTGGCGGTAACCACGGCGGTGGTGGTAACCATG 571

Qy     244 GGCAAGGCGATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAGATG 303
           |    ||||  ||||  | | |||  | |       | |    |  |||    | | ||
Db     572 GCGGCGGCGGCAACCACGGTGGTGGCGGCAACCATGGCGGTCGCGGCAACCATGGCGGGTG 631

Qy     304 GCCGGCAGCATTGTTGGTGCGATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTGGTG 363
           || |  | ||  || |  |||      | | ||  | | |        ||   | | ||
Db     632 GCGGCAACCACGGTGGTCGCGGCAACCACGGCGGTGGTGACAACCATGGCGGTGGCGGCA 691

Qy     364 GTCGCAGTTGTCGGGAAAGGCGCGGCGGCGAAACTGGGTAACGCGCTGAGCAAAATGATG 423
             | || || | |  |  | ||||| |    |    |  | | ||     ||  |||
Db     692 ACCACCGTGGTGGCGGCAACCATGGCGGTGGCGGTAACCACGGTGGTGGCGGCAACCATG 751

Qy     424 GGCGAAACGATTAAGAAGTTGGTGCC 449
           |  |   ||   |  | | |||||
Db     752 GCGGTGGCGGCAACCACGGTGGTGGC 777


RESULT   7
Q67401/c
ID   Q67401 standard; DNA; 10366 BP.
AC   Q67401;
DT   13-APR-1995  (first entry)
DE   Huntingtin DNA\IT15 gene.
KW   Polymerase chain reaction; primer; PCR; amplify; Huntingtin; IT15;
KW   open reading frame; polymorphic; (CAG)n; trinucleotide repeat;
KW   allele; Huntingdon's disease; HD; chromosome; juvenile HD; unstable;
KW   expandable; linkage disequilibrium; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             316. .9750
FT                   /*tag=  a
PN   EP-614977-A.
PD   14-SEP-1994.
PF   07-MAR-1994; 301587.
PR   05-MAR-1993; US-027498.
PR   01-JUL-1993; US-085000.
PA   (GEHO ) GEN HOSPITAL CORP.
PI   Ambrose CM,  Duyao MP,  Gusella JF,  MacDonald ME;
DR   WPI; 94-281205/35.
DR   P-PSDB; R58777.
PT   New huntingtin protein and related nucleic acid, antibodies etc.
PT   - for treatment and diagnosis of neuro-degenerative disease,
PT   specifically Huntington's disease
PS   Claim 5; Page 22-33; 66pp; English.
CC   This sequence represents a large gene, termed "Huntingtin" or
CC   "IT15" which is present in the proximal part of the 0.5 mb segment
CC   between members D4S180 and D4S182 on chromosome 4.  This gene
CC   spans about 210 kb and encodes a protein of approx. 348 kD.  The
CC   huntingtin open reading frame contains a polymorphic (CAG)n
CC   trinucleotide repeat with at least 17 alleles in the normal
CC   population, varying from about 11 to about 34 CAG copies.  On
CC   Huntingdon's disease (HD) chromosome, the length of the CAG repeat
CC   is substantially increased, with between 37 to at least 73 copies.
CC   This shows an apparent correlation with age of onset, the longest
CC   segments are detected in juvenile HD cases.  The presence of an
CC   unstable, expandable trinucleotide repeat on HD chromosomes in the
CC   region of the strongest linkage disequilibrium with the disorder
CC   suggest that this alteration underlies the dominant phenotype of HD.
SQ   Sequence   10366 BP;    2426 A;    2808 C;    2742 G;    2390 T;


  Query Match             4.4%;  Score 38.2;  DB 1;  Length 10366;
  Best Local Similarity   51.6%;  Pred. No. 0.32;
  Matches  111;  Conservative    0;  Mismatches  103;  Indels    1;  Gaps    1;

Qy      74 GCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTG 133
           ||   ||||| || |       | |  ||||||  |  |||   ||| || | ||    |
Db     552 GCCGGGTGGCGGCGGGGGCGGCGGCGGGGGCGGCTGCGGCTGAGGCAGCAGCGGCTGTGC 493

Qy     134 CGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCGTTTAT-TCAGCAG 192
           | |    || ||| || |||  || | ||||||| |||  | | ||    |  | || |
Db     492 CTGCGGCGGCGGCTGAGGAAGCTGAGGAGGCGGCGGCGGCGGCGGCGGTGGCGGCTG 433

Qy     193 GCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTGATTGGCAAGGCG 252
             |||    | | |  ||  ||   |||||  || | |  ||  |||| |       |  |
Db     432 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 373

Qy     253 ATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAA 287
           |      |||  || || ||| | |  | |  | |
Db     372 CTGCTGGAAGGACTTGAGGGACTCGAAGGCCTTCA 338


RESULT   8
T97924/c
ID   T97924 standard; DNA; 10366 BP.
AC   T97924;
DT   13-MAR-1998  (first entry)
DE   Composite sequence of a novel huntingtin (IT15) gene.
KW   Huntingin gene; IT15 gene; Huntington's disease; trinucleotide repeat;
KW   neurodegenerative disorder; HD; gene therapy; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             316. .9750
FT                   /*tag=  a
PN   US5686288-A.
PD   11-NOV-1997.
PF   20-MAY-1994; 246982.
PR   05-MAR-1993; US-027498.
PR   01-JUL-1993; US-085000.
PA   (GEHO ) GEN HOSPITAL CORP.
PI   Ambrose CM,  Duyao MP,  Gusella JF,  MacDonald ME;
DR   WPI; 97-558144/51.
DR   P-PSDB; W36887.
PT   Nucleic acid encoding huntingtin protein - useful for gene therapy
PT   of Huntington's disease
PS   Claim 3; Fig 4; 112pp; English.
CC   The present sequence represents a novel gene, termed huntingin or IT15.
CC   This gene encodes a previously undescribed protein. The huntingtin
CC   reading frame contains a polymorphic (CAG)n trinucleotide repeat with at
CC   least 17 alleles in the normal population, varying from about 11 to 34
CC   CAG copies. Huntington's disease (HD) is a progressive neurodegenerative
CC   disorder characterised by motor disturbance, cognitive loss and
CC   psychiatric manifestations. The genetic defect causing HD is assigned to
CC   chromosome 4. On HD chromosomes, the length of the trinucleotide CAG
CC   repeat is substantially increased, e.g. about 37 to at least 73 copies.
CC   The present nucleic acid and proteins encoded by it, may be used for the
CC   diagnosis or treatment of Huntington's disease. The huntingtin gene
CC   is especially used in gene therapy of a symptomatic or presymptomatic
CC   patient. The method comprises providing a functional huntingtin gene with
CC   a (CAG)n repeat of the normal range of 11-34 copies, or an antisense
```

```
CC   sequence, to the desired cells of the patient, in a manner that permits
CC   the expression of the huntingtin protein provided by the gene, or
CC   inhibits expression of the mutated huntingtin gene, for a time and in a
CC   quantity sufficient to provide the huntingtin function to the cells of
CC   the patient.
SQ   Sequence   10366 BP;    2426 A;    2809 C;    2742 G;    2389 T;


  Query Match             4.4%;  Score 38.2;  DB 1;  Length 10366;
  Best Local Similarity   51.6%;  Pred. No. 0.32;
  Matches  111;  Conservative    0;  Mismatches  103;  Indels    1;  Gaps    1;

Qy      74 GCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTG 133
           ||   ||||| || |      | |  ||||||  |  |||   ||| || | ||   |
Db     552 GCCGGGTGGCGGCGGGGGCGGCGGCGGGGGCGGCTGCGGCTGAGGCAGCAGCGGCTGTGC 493

Qy     134 CGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCGTTTAT-TCAGCAG 192
           | |    || ||| || |||  || | ||||||||  |||  | | ||     |  | || |
Db     492 CTGCGGCGGCGGCTGAGGAAGCTGAGGAGGCGGCGGCGGCGGCGGCGGCGGTGGCGGCTG 433

Qy     193 GCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTGATTGGCAAGGCG 252
             |||    | | |  ||  ||   |||||  || | |  ||  |||| |       |  |
Db     432 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 373

Qy     253 ATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAA 287
             |     |||  || || ||| | |  | |  | |
Db     372 CTGCTGGAAGGACTTGAGGGACTCGAAGGCCTTCA 338


RESULT   9
V05828/c
ID   V05828 standard; DNA; 10366 BP.
AC   V05828;
DT   01-JUN-1998  (first entry)
DE   Human huntingtin coding sequence.
KW   Human; huntingtin gene; Huntington's disease; chromosome; marker;
KW   locus; antisense; gene therapy; diagnosis; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             316. .9750
FT                   /*tag=  a
FT                   /product= huntingtin_protein
PN   US5693757-A.
PD   02-DEC-1997.
PF   30-MAY-1995; 453265.
PR   20-MAY-1994; US-246982.
PR   05-MAR-1993; US-027498.
PR   01-JUL-1993; US-085000.
PR   30-MAY-1995; US-453265.
PA   (GEHO ) GEN HOSPITAL CORP.
PI   Ambrose CM, Duyao MP, Gusella JF, MacDonald ME;
DR   WPI; 98-031815/03.
DR   P-PSDB; W44742.
PT   Huntingtin protein and related nucleic acid - for diagnosis or
PT   therapy of Huntington's disease
PS   Example 1; Fig 4; 112pp; English.
CC   This is the nucleotide sequence which encodes the human huntingtin
CC   protein.  The gene sequence is characterised in that it contains a
CC   number of CAG repeats in the 5' region (in this case 23 repeats).
CC   In healthy individuals the usual number of CAG repeats in the gene
CC   is 11-34, whereas in patients suffering from Huntington's disease (HD),
CC   the number of repeats increases to 37-73 or 37-86.  The huntingtin gene
CC   spans 210 kb and encodes a protein of 348 kD. The gene is found in a
CC   500 kb region between the chromosomal markers D4S180 and D4S182 and is
CC   preferentially mapped to the locus 4p16.3.  The protein or the gene
CC   encoding it, is useful for detecting a predisposition to develop HD,
CC   for diagnosis and treatment of HD, especially by antisense and gene
CC   therapy.
SQ   Sequence   10366 BP;    2426 A;    2809 C;    2742 G;    2389 T;


  Query Match             4.4%;  Score 38.2;  DB 1;  Length 10366;
  Best Local Similarity   51.6%;  Pred. No. 0.32;
  Matches  111;  Conservative    0;  Mismatches  103;  Indels    1;  Gaps    1;

Qy      74 GCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTG 133
           ||   ||||| || |      | |  ||||||  |  |||   ||| || | ||   |
Db     552 GCCGGGTGGCGGCGGGGGCGGCGGCGGGGGCGGCTGCGGCTGAGGCAGCAGCGGCTGTGC 493

Qy     134 CGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCGTTTAT-TCAGCAG 192
           | |    || ||| || |||  || | ||||||||  |||  | | ||     |  | || |
Db     492 CTGCGGCGGCGGCTGAGGAAGCTGAGGAGGCGGCGGCGGCGGCGGCGGCGGTGGCGGCTG 433

Qy     193 GCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTGATTGGCAAGGCG 252
             |||    | | |  ||  ||   |||||  || | |  ||  |||| |       |  |
Db     432 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 373

Qy     253 ATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAA 287
             |     |||  || || ||| | |  | |  | |
Db     372 CTGCTGGAAGGACTTGAGGGACTCGAAGGCCTTCA 338


RESULT  10
X18930/c
ID   X18930 standard; DNA; 515 BP.
AC   X18930;
DT   11-MAY-1999  (first entry)
DE   Human huntingtin gene.
KW   Huntingtin protein; osm-10 gene; regulatory region; TREND disease;
KW   Huntington's disease; triplet repeat neurological disease;
KW   transgenic nematode; ss.
OS   Homo sapiens.
PN   WO9902652-A1.
PD   21-JAN-1999.
PF   10-JUL-1998; U14376.
PR   11-JUL-1997; US-052324.
PA   (GEHO ) GEN HOSPITAL CORP.
PI   Faber P, Hart AC, MacDonald M;
DR   WPI; 99-120856/10.
DR   P-PSDB; W99022.
PT   Transgenic nematode containing nucleic acid related to
PT   triplet-repeat neurological disease - useful in screening for agents
PT   that can treat or prevent such diseases, especially Huntington's
PT   disease
PS   Claim 3; Fig 2; 40pp; English.
CC   The present invention describes a transgenic nematode (A) comprising a
CC   nucleic acid (I), associated with a triplet-repeat neurological disease
CC   (TREND), linked to a neuron-specific regulator. Also described are: (1)
CC   transgenic nematode (A') exhibiting progressive neuronal cell death; (2)
CC   expressing a nucleic acid in neuron-specific manner by linking it to an
CC   osm-10 regulator, in a region that, in the natural osm-10 gene, is
CC   occupied by all or part of the osm-10 coding sequence; and (3) transgene
CC   comprising (I) linked to the osm-10 regulator as specified in (2). (A)
CC   and (A') are used to screen for compounds useful in treatment or
CC   prevention of TRENDs (e.g. Huntington's disease, spinocerebellar ataxia,
CC   dentatorubral-pallidoluysian atrophy and spinal bulbar muscular
CC   dystrophy, also non-CAG related disorders). The nematodes are accurate
CC   models of TRENDs, i.e. they show the progressive nerve cell dysfunction
CC   characteristic of human disease and this can be monitored over time,
CC   using a population of animals that hatched together. Since they begin to
CC   show neuronal cell loss as early as day 7 from hatching, the nematodes
CC   can be used to test palliative treatments and the aetiology of TRENDs
CC   may also be studied. Nematodes have short generation times; large
CC   numbers can be manipulated easily, allowing accurate staging of many of
CC   them for high throughput testing, and can be bred to produce identical
CC   offspring. The osm-10 promoter is active only in neuronal tissue. The
CC   present sequence represents the human huntingtin gene from the present
CC   invention.
SQ   Sequence   515 BP;    125 A;    169 C;    133 G;    88 T;


  Query Match             4.4%;  Score 38.2;  DB 1;  Length 515;
  Best Local Similarity   51.6%;  Pred. No. 0.086;
  Matches  111;  Conservative    0;  Mismatches  103;  Indels    1;  Gaps    1;

Qy      74 GCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTG 133
```

```
Db    239 GCCGGGTGGCGGCGGGGGCGGCGGCGGGGGCGGCTGCGGCTGAGGCAGCAGCGGCTGTGC 180

Qy    134 CGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCGTTTAT-TCAGCAG 192

Db    179 CTGCGGCGGCGGCTGAGGAAGCTGAGGAGGCGGCGGCGGCGGCGGCGGTGGCGGCTG 120

Qy    193 GCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTGATTGGCAAGGCG 252

Db    119 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 60

Qy    253 ATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAA 287

Db     59 CTGCTGGAAGGACTTGAGGGACTCGAAGGCCTTCA 25


RESULT 11
Q31795/c
ID   Q31795 standard; cDNA; 342 BP.
AC   Q31795;
DT   21-APR-1993  (first entry)
DE   PRP2 gene.
KW   Proline-rich protein; PRP1; PRP2; PRP3; PRP4; PRP3g12; pistil;
KW   flower; sterile; plant; hydroxyproline; glycoprotein; gum; ss.
OS   Nicotiana alata.
FH   Key             Location/Qualifiers
FT   cds             1. .342
FT                   /*tag= a
FT                   /label= PRP2
FT   repeat_region   58. .210
FT                   /*tag= b
FT                   /note= "see CC"
FT   repeat_region   220. .327
FT                   /*tag= c
FT                   /note= "see CC"
PN   WO9220713-A.
PD   26-NOV-1992.
PF   15-MAY-1992; AU0222.
PR   15-MAY-1991; AU-006159.
PA   (UYME ) UNIV MELBOURNE.
PI   Chen C, Clarke AE;
DR   WPI; 92-415710/50.
DR   P-PSDB; R29163.
PT   Isolated DNA sequence encoding proline rich proteins of Nicotiana
PT   alata - used to express useful proteins and identify regulatory
PT   elements for pistil specific expression, to give e.g. longer
PT   flower life
PS   Claim 2; Pag 47 + Fig 2; 57pp; English.
CC   The sequence below is the 342 bp insert from clone PRP2.1.
CC   The incomplete polypeptide encoded by this cDNA contains 49% proline
CC   residues in total of 114 amino acids.  It can be divided into two
CC   domains, one (encoded by nucleotides 58-210) contains 5 repeats of
CC   (Pro)4-Ala interspersed with 4 repeats of the pentapeptide Gln-Leu-
CC   Pro-Ile-Arg.  These repeat sequences in turn make up two copies
CC   of a higher order repetitive unit of 23 amino acids (Pro)4-Ala-Gln-
CC   Leu-Pro-Ile-Arg-Gln-(Pro)4-Ala-Thr-Gln-Leu-Pro-Ile-Arg-Lys.
CC   The other domain (nucleotides 220-327) is composed mainly of
CC   (Pro)2-5-Gly-Tyr repeats (7) that are tandemly reiterated.  At the
CC   nucleotide level, the sequences encoding the various repeats are
CC   highly conserved.
CC   The sequences of the PRP1, PRP2, PRP3 (cDNA and genomic DNA),
CC   PRP4 genes are given in Q31794-98.  The DNA sequences may be used to
CC   identify regulatory elements responsible for pistil-specific expression.
CC   Such elements may be used to provide pistil-specific expression of
CC   foreign or heterologous genes, e.g. to enable longer flower life or
CC   in the creation of male- or female-sterile plants.  The PRPs may be
CC   converted by hydroxylation and glycosylation into hydroxyproline-rich
CC   glycoproteins.  These could by used as effective substitutes for gums,
CC   e.g. guar gum and gum arabic, in foods and for non-food applications,
CC   where emulsification, thickening and stabilisation are required.
SQ   Sequence   342 BP;    105 A;    139 C;    53 G;    45 T;

  Query Match             4.4%;  Score 38;  DB 1;  Length 342;
  Best Local Similarity   45.4%;  Pred. No. 0.081;
  Matches  137;  Conservative    0;  Mismatches  165;  Indels    0;  Gaps    0;

Qy    123 GGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCGTT 182

Db    305 GGCGGCGGATACCCTGGCGGTGGTGGTGGATACCCTGGTGGCGGATATCCTGGTGGCGGA 246

Qy    183 TATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTGAT 242

Db    245 TATCCTGGTGGCGGATATCCCGGTGGTTGAGTATATGCTGGTGGTGGTGGTTTCCTTATT 186

Qy    243 TGGCAAGGCGATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAGAT 302

Db    185 GGTAGCTGAGTAGCTGGTGGTGGTGGTTGCCTTATTGGTAACTGAGCTGGTGGTGGTGGT 126

Qy    303 GGCCGGCAGCATTGTTGGTGCGATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTGGT 362

Db    125 TTCCTTATTGGTAGCTGAGTAGCTGGTGGTGGTGGTTGCCTTATTGGTAACTGAGCTGGT 66

Qy    363 GGTCGCAGTTGTCGGGAAAGGCGCGGCGGCGAAACTGGGTAACGCGCTGAGCAAAATGAT 422

Db     65 GGTGGTGGTAGTGGTAATTTTGGTGGTGGTGCTCGTTGCATTGGCGGTGGCTGAGAAGGT 6

Qy    423 GG 424

Db      5 GG 4


RESULT 12
V20700/c
ID   V20700 standard; DNA; 5163 BP.
AC   V20700;
DT   17-AUG-1998  (first entry)
DE   Cryptosporidium parvum GP900 antigen open reading frame.
KW   Open reading frame; ORF; antigen; GP900; cryptosporidium; infection;
KW   antibody; prophylaxis; treatment; inhibition; retardation;
KW   detection; diagnosis; human; ds.
OS   Cryptosporidium parvum.
FH   Key             Location/Qualifiers
FT   CDS             1. .5163
FT                   /*tag= a
FT                   /product= GP900 antigen
FT   misc_feature    524. .1270
FT                   /*tag= b
FT                   /note= "region containing NINC mutations"
PN   WO9806430-A1.
PD   19-FEB-1998.
PF   11-AUG-1997; U14104.
PR   14-AUG-1996; US-700651.
PA   (REGC ) UNIV CALIFORNIA.
PI   Gut J, Leech J, Nelson RC, Petersen C;
DR   WPI; 98-159290/14.
DR   P-PSDB; W48299.
PT   Anti-Cryptosporidium antibody - used to develop products for
PT   detection, diagnosis, prophylaxis or treatment of Cryptosporidium
PT   infections
PS   Claim 32; Pages 60-62; 89pp; English.
CC   The sequence is that encoding the GP900 antigen which may be used
CC   in the production of anti-Cryptosporidium antibodies. These can be
CC   used for the prophylaxis, treatment, inhibition or retardation of
CC   a Cryptosporidium infection in humans or in animals such as calves.
CC   They can also be used for the detection and diagnosis of related
CC   infections.
SQ   Sequence   5163 BP;    1875 A;    1137 C;    873 G;    1278 T;

  Query Match             4.3%;  Score 37.8;  DB 1;  Length 5163;
  Best Local Similarity   44.9%;  Pred. No. 0.31;
  Matches  144;  Conservative    0;  Mismatches  177;  Indels    0;  Gaps    0;

Qy     68 TTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCTGCGGTGG 127
```

```
Db    980 TTGTGGTGGTTGTAGTAGTAGTTGTTGTTGTTGTTGTTGTTGTAGTTGTTGTTGTTG 921

Qy    128 GACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCGTTTATTC 187

Db    920 GTTTCTTGGTAGTTGTGGTTGTGGTTGTGGTTGTTGTGGTTGTAGTTGTTGTGGTTGTAG 861

Qy    188 AGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTGATTGGCA 247

Db    860 TTGTGGTTGTAGTTGTTGTGGTTGTTGTTGTTGTGGTTGTAGTTGTTGTGGTTGTAGTTG 801

Qy    248 AGGCGATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAGATGGCCG 307

Db    800 TGGTTGTAGTTGTTGTGGTTGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAG 741

Qy    308 GCAGCATTGTTGGTGCGATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTGGTGGTCG 367

Db    740 TAGTCGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 681

Qy    368 CAGTTGTCGGGAAAGGCGCGG 388

Db    680 TTGTTGTCGTAGTAGTAGTGG 660
```

RESULT 13
V38816/c
ID V38816 standard; DNA; 1107 BP.
AC V38816;
DT 09-OCT-1998 (first entry)
DE DNA encoding a durum wheat glutenin protein.
KW Glutenin gene; durum wheat; low-molecular-weight;
KW transgenic durum wheat; ss.
OS Triticum durum.
PN FR2757538-A1.
PD 26-JUN-1998.
PF 18-DEC-1997; 016059.
PR 19-DEC-1996; IT-MI2663.
PA (ITUY-) ITAL MIN UNIV RICERCA SCI & TECNOLOGICA.
PI Cardelli LE, DOvidio R, Marchitelli C, Porceddu E;
DR WPI; 98-365055/32.
DR P-PSDB; W62647.
PT Durum wheat glutenin gene - coding for glutenin protein of low
PT molecular weight
PS Claim 1; Page 13; 18pp; French.
CC The present sequence represents a glutenin gene, and is isolated from
CC the genomic DNA of Triticum durum L. The gene codes for a
CC low-molecular-weight glutenin protein and can be used to produce
CC transgenic durum wheat plants with "better quality characteristics"
CC [no details given].
SQ Sequence 1107 BP; 368 A; 377 C; 152 G; 210 T;

Query Match 4.3%; Score 37.4; DB 1; Length 1107;
Best Local Similarity 46.4%; Pred. No. 0.2;
Matches 162; Conservative 0; Mismatches 181; Indels 6; Gaps 1;

```
Qy     25 ATTATGGGATGTATCGGGAAAGTCCTCGGCGCGCTGCTAACCATTGTCAGCGTTGTGGCC 84

Db    463 ATGGTGGTAGTTGCTGTGAAAATGGTGGTGGTTGTTGCTGCGAAAATGGTGGTTGTTGCG 404

Qy     85 GCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTG 144

Db    403 GTAGTACTGGTTGTTGCTGCGAAAATGGTGGTAGTTGCTGTGAAAATGGTGGTAGTTGTT 344

Qy    145 GCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCGTTTATTCAGCAGGCGCTAAACCCG 204

Db    343 GCTGCGAAAATGATGGTTGTTGCGGTAGAACTGGTTGTTGTT------GCTGTGAAAATG 290

Qy    205 ATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTGATTGGCAAGGCGATTACCAAAGCG 264

Db    289 GTGGTTGTTGTTGCTGCGAAAATGGTGGTAGTTGTTGCTGCGAAAATGATGGTTGTTGCG 230

Qy    265 CTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAGATGGCCGGCAGCATTGTTGGTGCG 324
```

```
Db    229 GTAGAACTGGTTGTTGTTGCTGTGAAAATGGTGGTTGTTGTTGCTGCGAAAATGGTGGTT 170

Qy    325 ATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTGGTGGTCGCAGTTG 373

Db    169 GTTGTTGTTGCGATAATGGTGGTTGTTGTTGTTGCTGTGAACATGGTTG 121
```

RESULT 14
Q03877
ID Q03877 standard; DNA; 864 BP.
AC Q03877;
DT 31-AUG-1989 (first entry)
DE Merozoite surface antigen-2 (MSA2) Indochina I isolate gene
KW P.falciparum merozoite surface antigen; malaria; ss.
OS Plasmodium falciparum.
FH Key Location/Qualifiers
FT cds 1. .859
FT /*tag= a
FT /label=MSA2 Indochina I isolate
FT repeat_region 169. .312
FT /*tag= b
FT /rpt_type=tandem
FT repeat_region 379. .402
FT /*tag= c
FT /rpt_unit=t
PN WO9002752-A.
PD 22-MAR-1990.
PF 12-SEP-1989; AU00388.
PR 24-AUG-1989; AU-005962, AU-000382.
PA (SARA-) Saramane Pty Ltd.
PI Smythe JA, Anders RF, Kemp DJ, Coppel RL, Saul AJ, Jones GL, Irving DO,
PI Dyer SL, Chee DW;
DR WPI; 90-115951/15.
DR P-PSDB; R05877.
PT Allelic variants of P.falciparum merozoite surface antigen-used in prodn.
PT of antibodies and in vaccines against malaria.
PS Disclosure; fig 2; 53pp; English.
CC This sequence encodes the merozoite surface antigen-2 (MSA2) Indochina I
CC isolate of plasmodium falciparum which is used to generate Abs that
CC inhibit the growth of P.falciparum. The antigenic protein is hence useful
CC in immunisation against malaria. Indochina isolate differs from others
CC eg 3D7 and hence reacts with a specific repeat probe based on its
CC sequence. The isolates do have conserved regions however, located at the
CC N- and C-termini.
CC See also Q03876, Q03878, Q03879, Q03880, Q03881, Q03882, Q03884, R03974,
CC R03975, R03976, R03977, R03978, R03979, R03980 and R03985.
SQ Sequence 864 BP; 288 A; 157 C; 179 G; 240 T;

Query Match 4.3%; Score 37; DB 1; Length 864;
Best Local Similarity 47.6%; Pred. No. 0.24;
Matches 109; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

```
Qy     47 TCCTCGGCGCGCTGCTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGA 106

Db    144 TCCTACTGGTGCTAGTGGTAGTGCTGGTGGTAGTGCTGGTGGTAGTGCTGGTGGTAGTGC 203

Qy    107 GTCTGGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCGG 166

Db    204 TGGTGGTAGTGCTGGTGGTAGTGCTGGTGGTAGTGCTGGTGGTAGTGCTGGTGGTAGTGC 263

Qy    167 CGACGGGAGTGTCGTTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAGC 226

Db    264 TGGTGGTAGTGCTGGTGGTAGTGCTGGTGGTAGTGCTGGTGGTAGTGCTGGTTCTGGTGA 323

Qy    227 CGTTAATGGAGCTGATTGGCAAGGCGATTACCAAAGCGCTGGAAGGATT 275

Db    324 TGGTAATGGTGCTAATCCTGGTGCAGATGCTGAGAGAAGTCCAAGTACT 372
```

RESULT 15
V20701/c
ID V20701 standard; DNA; 5318 BP.

```
AC   V20701;
DT   17-AUG-1998  (first entry)
DE   Cryptosporidium parvum GP900 antigen open reading frame and 3' region.
KW   Open reading frame; ORF; antigen; GP900; cryptosporidium; infection;
KW   antibody; prophylaxis; treatment; inhibition; retardation;
KW   detection; diagnosis; human; 3' region; ds.
OS   Cryptosporidium parvum.
FH   Key             Location/Qualifiers
FT   CDS             1. .5166
FT                   /*tag= a
FT                   /product= GP900 antigen
FT   misc_feature    524. .1270
FT                   /*tag= b
FT                   /note= "region containing NINC mutations"
PN   WO9806430-A1.
PD   19-FEB-1998.
PF   11-AUG-1997; U14104.
PR   14-AUG-1996; US-700651.
PA   (REGC ) UNIV CALIFORNIA.
PI   Gut J, Leech J, Nelson RC, Petersen C;
DR   WPI; 98-159290/14.
DR   P-PSDB; W48299.
PT   Anti-Cryptosporidium antibody - used to develop products for
PT   detection, diagnosis, prophylaxis or treatment of Cryptosporidium
PT   infections
PS   Claim 32; Pages 62-63; 89pp; English.
CC   The sequence is that encoding the GP900 antigen which may be used
CC   in the production of anti-Cryptosporidium antibodies. These can be
CC   used for the prophylaxis, treatment, inhibition or retardation of
CC   a Cryptosporidium infection in humans or in animals such as calves.
CC   They can also be used for the detection and diagnosis of related
CC   infections.
SQ   Sequence   5318 BP;    1944 A;    1153 C;    888 G;    1333 T;


  Query Match              4.2%;  Score 36.2;  DB 1;  Length 5318;
  Best Local Similarity   44.5%;  Pred. No. 0.9;
  Matches  143;  Conservative    0;  Mismatches  178;  Indels    0;  Gaps    0;

Qy      68 TTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCTGCGGTGG 127
           ||||    ||||| |  |  ||| ||   | ||  |    | | |   | | ||  || |
Db     980 TTGTGGTGGTTGTAGTAGTAGTTGTTGTTGTTGTTGTTGTTGTAGTTGTTGTTGTTG 921

Qy     128 GACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCGTTTATTC 187
           |  |   |||| | ||||  |  |   | ||    | ||     |  ||   | || |
Db     920 GTTTCTTGGTAGTTGTGGTTGTGGTTGTGGTTGTTGTGGTTGTAGTTGTTGTGGTTGTAG 861

Qy     188 AGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTGATTGGCA 247
             ||   ||       |  | |    |||  ||       |  | | | |  | |
Db     860 TTGTGGTTGTAGTTGTTGTGGTTGTTGTTGTTGTGGTTGTAGTTGTTGTGGTTGTAGTTG 801

Qy     248 AGGCGATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAGATGGCCG 307
            ||   |      | | | | || | |||  || |      |   | ||   | |  |
Db     800 TGGTTGTAGTTGTTGTGGTTGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAG 741

Qy     308 GCAGCATTGTTGGTGCGATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTGGTGGTCG 367
           | |||||| ||   |||| |  | | |||   | || |  ||  ||||  | || |
Db     740 TAGTCGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTTTTGTTG 681

Qy     368 CAGTTGTCGGGAAAGGCGCGG 388
             |||||||    ||  | ||
Db     680 TTGTTGTCGTAGTAGTAGTGG 660


Search completed: October  5, 1999, 17:58:06
Job time: 11328 sec
```

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 1999, 14:49:17 ; Search time 2414.69 Seconds
(without alignments)
1145.849 Million cell updates/sec

Title: US-09-068-804-1
Perfect score: 870
Sequence: 1 CGCAAAGCCGAGGAAACGAA.....TGCGCCAGAGTCGCGCATAA 870

Scoring table: IDENTITY_NUC

Searched: 679419 seqs, 1590154680 residues

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_st:*
14: gb_sts:*
15: gb_sy:*
16: gb_un:*
17: gb_vi:*
18: em_fun:*
19: em_htg:*
20: em_hum1:*
21: em_hum2:*
22: em_in:*
23: em_om:*
24: em_or:*
25: em_ov:*
26: em_pat:*
27: em_ph:*
28: em_pl:*
29: em_ro:*
30: em_sts:*
31: em_sy:*
32: em_un:*
33: em_vi:*
34: gb_htg1:*
35: gb_htg2:*
36: gb_in1:*
37: gb_in2:*
38: em_ba1:*
39: em_ba2:*
40: em_hum3:*
41: em_hum4:*
42: gb_pr4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	870	100.0	3622	1	STU25631	U25631 Salmonella
2	870	100.0	3441	2	STU30491	U30491 Salmonella
3	865.2	99.4	2995	1	SDU66877	U66877 Salmonella
4	858.8	98.7	5393	1	STSIPGNS	X82670 S.typhi spa
5	178.6	20.5	6869	1	SDIPABCD	X60777 Shigella dy
6	178.6	20.5	7663	1	SFPVR2	X15319 Shigella fl
7	178.6	20.5	5255	1	SHFIPABCD	J04117 S.flexneri
8	178.6	20.5	4761	1	W10PPIBCA	M34849 Plasmid pWR
9	72.6	8.3	207466	35	AC007249	AC007249 Homo sapi
10	72	8.3	158880	34	AC007009	AC007009 Homo sapi
c 11	71.4	8.2	137137	42	AC004928	AC004928 Homo sapi
12	70.2	8.1	168844	35	AC007263	AC007263 Homo sapi
c 13	67.8	7.8	162269	11	AC000097	AC000097 Homo sapi
c 14	67.8	7.8	165321	35	AC006549	AC006549 Homo sapi
c 15	67.8	7.8	179269	42	AC006547	AC006547 , complet
16	67.6	7.8	171752	35	AC007381	AC007381 Homo sapi
c 17	67.4	7.7	141077	35	AC005918	AC005918 Homo sapi
18	65.4	7.5	858	10	HSHZREP42	X59828 Human chrom
19	62.8	7.2	103826	34	HSB11L7C8	AL078472 Homo sapi
c 20	61.6	7.1	41185	10	HSB1L1C6	AL049849 Homo sapi
21	59.6	6.9	131453	34	HS989H11	Z83851 Homo sapien
22	58.4	6.7	256172	35	AC005139	AC005139 Plasmodiu
c 23	57.4	6.6	1988	11	AC002494	AC002494 Human PCR
24	56.8	6.5	27251	9	HS293L6C	Z83732 Human DNA s
25	56.8	6.5	99497	10	HS293L6	AL049749 Human DNA
c 26	55.8	6.4	175484	34	HS215D11	AL034417 Homo sapi
c 27	55.2	6.3	284972	35	AC005505	AC005505 Plasmodiu
c 28	55	6.3	143701	9	HS29C18	Z97192 Homo sapien
29	54.6	6.3	174428	11	AC005821	AC005821 Homo sapi
c 30	54.2	6.2	131427	9	HS37J18	Z98747 Human DNA s
c 31	54.2	6.2	175609	11	AC005243	AC005243 Homo sapi
c 32	54	6.2	7218	5	I66494	I66494 Sequence 14
33	54	6.2	112460	34	HS1104E15	AL022312 Homo sapi
34	53	6.1	155074	11	AC003693	AC003693 Human Chr
35	53	6.1	105892	34	HSY214H10_4	Continuation (5 of
c 36	52.6	6.0	135258	9	HS127B20	Z83838 Human DNA s
37	51.4	5.9	834	7	OSGRP09MR	X54449 O.sativa GR
38	51.4	5.9	3250	7	OSU40708	U40708 Oryza sativ
c 39	51.2	5.9	117338	10	HS173D1	AL031984 Human DNA
40	50.2	5.8	888	7	NTGRPR	X74106 N.tabacum g
41	49.8	5.7	1925	7	PVGRP18	X13596 Bean DNA fo
42	49.8	5.7	206344	35	AC007240	AC007240 Homo sapi
c 43	49.8	5.7	242867	35	AC007623	AC007623 Homo sapi
44	49	5.6	103270	7	ATF1N20	AL022140 Arabidops
45	48.6	5.6	86801	8	AC007190	AC007190 Genomic s

ALIGNMENTS

RESULT 1
STU25631
LOCUS STU25631 3622 bp DNA BCT 06-SEP-1995
DEFINITION Salmonella typhimurium invasin homologs SicA (sicA), SipB (sipB) and SipC (sipC) genes, complete cds.
ACCESSION U25631
NID g975293
VERSION U25631.1 GI:975293
KEYWORDS .
SOURCE Salmonella typhimurium.
ORGANISM Salmonella typhimurium
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Salmonella.
REFERENCE 1 (bases 1 to 3622)
AUTHORS Kaniga,K., Tucker,S., Trollinger,D. and Galan,J.E.
TITLE Homologs of the Shigella IpaB and IpaC invasins are required for Salmonella typhimurium entry into cultured epithelial cells
JOURNAL J. Bacteriol. 177 (14), 3965-3971 (1995)
MEDLINE 95332200
REFERENCE 2 (bases 1 to 3622)
AUTHORS Galan,J.E.
TITLE Direct Submission
JOURNAL Submitted (23-APR-1995) Jorge E. Galan, Molecular Genetics and Microbiology, School of Medicine, SUNY Stony Brook, Stony Brook, NY

```
          11794-5222, USA
FEATURES             Location/Qualifiers
     source          1. .3622
                     /organism="Salmonella typhimurium"
                     /strain="SL1344"
                     /db_xref="taxon:602"
     gene            75. .572
                     /gene="sicA"
     CDS             75. .572
                     /gene="sicA"
                     /codon_start=1
                     /transl_table=11
                     /product="SicA"
                     /protein_id="AAA75168.1"
                     /db_xref="PID:g975294"
                     /db_xref="GI:975294"
                     /translation="MDYQNNVSEERVAEMIWDAVSEGATLKDVHGIPQDMMDGLYAHA
                     YEFYNQGRLDEAETFFRFLCIYDFYNPDYTMGLAAVCQLKKQFQKACDLYAVAFTLLK
                     NDYRPVFFTGQCQLLMRKAAKARQCFELVNERTEDESLRAKALVYLEALKTAETEQHS
                     EQEKE"
     gene            575. .2356
                     /gene="sipB"
     CDS             575. .2356
                     /gene="sipB"
                     /codon_start=1
                     /transl_table=11
                     /product="SipB"
                     /protein_id="AAA75169.1"
                     /db_xref="PID:g975295"
                     /db_xref="GI:975295"
                     /translation="MVNDASSISRSGYTQNPRLAEAAFEGVRKNTDFLKAADKAFKDV
                     VATKAGDLKAGTKSGESAINTVGLKPPTDAAREKLSSEGQLTLLLGKLMTLLGDVSLS
                     QLESRLAVWQAMIESQKEMGIQVSKEFQTALGEAQEATDLYEASIKKTDTAKSVYDAA
                     TKKLTQAQNKLQSLDPADPGYAQAEAAVEQAGKEATEAKEALDKATDATVKAGTDAKA
                     KAEKADNILTKFQGTANAASQNQVSQGEQDNLSNVARLTMLMAMFIEIVGKNTEESLQ
                     NDLALFNALQEGRQAEMEKKSAEFQEETRKAEETNRIMGCIGKVLGALLTIVSVVAAV
                     FTGGASLALAAVGLAVMVADEIVKAATGVSFIQQALNPIMEHVLKPLMELIGKAITKA
                     LEGLGVDKKTAEMAGSIVGAIVAAIAMVAVIVVVAVVGKGAAAKLGNALSKMMGETIK
                     KLVPNVLKQLAQNGSKLFTQGMQRITSGLGNVGSKMGLQTNALSKELVGNTLNKVALG
                     MEVTNTAAQSAGGVAEGVFIKNASEALADFMLARFAMDQIQQWLKQSVEIFGENQKVT
                     AELQKAMSSAVQQNADASRFILRQSRA"
     gene            2384. .3613
                     /gene="sipC"
     CDS             2384. .3613
                     /gene="sipC"
                     /codon_start=1
                     /transl_table=11
                     /product="SipC"
                     /protein_id="AAA75170.1"
                     /db_xref="PID:g975296"
                     /db_xref="GI:975296"
                     /translation="MLISNVGINPAAYLNNHSVENSSQTASQSVSAKDILNSIGISSS
                     KVSDLGLSPTLSAPAPGVLTQTPGTITSSLKASIQNTDMNQDLNALANNVTTKANEVV
                     QTQLREQQAEVGKFFDISGMSSSAVALLAAANTLMLTLNQADSKLSGKLSLVSFDAAK
                     TTASSMMREGMNALSGSISQSALQLGITGVGAKLEYKGLQNERGALKHNAAKIDKLTT
                     ESHSIKNVLNGQNSVKLGAEGVDSLKSLNIRKPVPMRRKILMMRRLNLMPEPAPRKVW
                     VLKTVINKVSLNIYILSKRLESVESDIRLEQNYMDITRIDSAQDADDGRSDYEELGHG
                     RWYCRGVRAVRRYSGNVSEQQISQVNNRVASTASDEARESSRKSTSLIQEMLKTMESI
                     NQSKASALAAIAGNIRA"
BASE COUNT     1049 a    759 c   1008 g    806 t
ORIGIN


  Query Match           100.0%;  Score 870;  DB 1;  Length 3622;
  Best Local Similarity  100.0%;  Pred. No. 4.2e-203;
  Matches  870;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCAAAGCCGAGGAAACGAACCGCATTATGGGATGTATCGGGAAAGTCCTCGGCGCGCTG 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1487 CGCAAAGCCGAGGAAACGAACCGCATTATGGGATGTATCGGGAAAGTCCTCGGCGCGCTG 1546

Qy      61 CTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCT 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    1547 CTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCT 1606

Qy     121 GCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCG 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1607 GCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCG 1666

Qy     181 TTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTG 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1667 TTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTG 1726

Qy     241 ATTGGCAAGGCGATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAG 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1727 ATTGGCAAGGCGATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAG 1786

Qy     301 ATGGCCGGCAGCATTGTTGGTGCGATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTG 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1787 ATGGCCGGCAGCATTGTTGGTGCGATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTG 1846

Qy     361 GTGGTCGCAGTTGTCGGGAAAGGCGCGGCGGCGAAACTGGGTAACGCGCTGAGCAAAATG 420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1847 GTGGTCGCAGTTGTCGGGAAAGGCGCGGCGGCGAAACTGGGTAACGCGCTGAGCAAAATG 1906

Qy     421 ATGGGCGAAACGATTAAGAAGTTGGTGCCTAACGTGCTGAAACAGTTGGCGCAAAACGGC 480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1907 ATGGGCGAAACGATTAAGAAGTTGGTGCCTAACGTGCTGAAACAGTTGGCGCAAAACGGC 1966

Qy     481 AGCAAACTCTTTACCCAGGGGATGCAACGTATTACTAGCGGTCTGGGTAATGTGGGTAGC 540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1967 AGCAAACTCTTTACCCAGGGGATGCAACGTATTACTAGCGGTCTGGGTAATGTGGGTAGC 2026

Qy     541 AAGATGGGCCTGCAAACGAATGCCTTAAGTAAAGAGCTGGTAGGTAATACCCTAAATAAA 600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2027 AAGATGGGCCTGCAAACGAATGCCTTAAGTAAAGAGCTGGTAGGTAATACCCTAAATAAA 2086

Qy     601 GTGGCGTTGGGCATGGAAGTCACGAATACCGCAGCCCAGTCAGCCGGTGGTGTTGCCGAG 660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2087 GTGGCGTTGGGCATGGAAGTCACGAATACCGCAGCCCAGTCAGCCGGTGGTGTTGCCGAG 2146

Qy     661 GGCGTATTTATTAAAAATGCCAGCGAGGCGCTTGCTGATTTTATGCTCGCCCGTTTTGCC 720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2147 GGCGTATTTATTAAAAATGCCAGCGAGGCGCTTGCTGATTTTATGCTCGCCCGTTTTGCC 2206

Qy     721 ATGGATCAGATTCAGCAGTGGCTTAAACAATCCGTAGAAATATTTGGTGAAAACCAGAAG 780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2207 ATGGATCAGATTCAGCAGTGGCTTAAACAATCCGTAGAAATATTTGGTGAAAACCAGAAG 2266

Qy     781 GTAACGGCGGAACTGCAAAAAGCCATGTCTTCTGCGGTACAGCAAAATGCGGATGCTTCG 840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2267 GTAACGGCGGAACTGCAAAAAGCCATGTCTTCTGCGGTACAGCAAAATGCGGATGCTTCG 2326

Qy     841 CGTTTTATTCTGCGCCAGAGTCGCGCATAA 870
             ||||||||||||||||||||||||||||||
Db    2327 CGTTTTATTCTGCGCCAGAGTCGCGCATAA 2356


RESULT   2
STU30491
LOCUS       STU30491     3441 bp    DNA             BCT       19-JAN-1996
DEFINITION  Salmonella typhimurium SspA (sspA) and SspB (sspB) genes, partial
            cds, and SspC (sspC) and SspD (sspD) genes, complete cds.
ACCESSION   U30491
NID         g1117811
VERSION     U30491.1  GI:1117811
KEYWORDS    .
SOURCE      Salmonella typhimurium.
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Salmonella.
REFERENCE   1  (bases 1 to 3441)
  AUTHORS   Hueck,C.J., Hantman,M.J., Bajaj,V., Johnston,C., Lee,C.A. and
            Miller,S.I.
  TITLE     Salmonella typhimurium secreted invasion determinants are
            homologous to Shigella Ipa proteins
```

```
  JOURNAL   Mol. Microbiol. 18 (3), 479-490 (1995)
  MEDLINE   96342378
REFERENCE   2  (bases 1 to 3441)
  AUTHORS   Miller,S.I.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUN-1995) Samuel I. Miller, Infectious Disease Unit,
            Massachusetts General Hospital, Harvard Medical School, Fruit
            Street, Boston, MA 02114, USA
FEATURES             Location/Qualifiers
     source          1. .3441
                     /organism="Salmonella typhimurium"
                     /strain="SL1344"
                     /db_xref="taxon:602"
                     /map="59 min"
     gene            1. .870
                     /gene="sspB"
     CDS             <1. .870
                     /gene="sspB"
                     /note="secreted protein"
                     /codon_start=1
                     /transl_table=11
                     /function="invasion determinant"
                     /product="SspB"
                     /protein_id="AAC43546.1"
                     /db_xref="PID:g1117812"
                     /db_xref="GI:1117812"
                     /translation="RKAEETNRIMGCIGKVLGALLTIVSVVAAVFTGGASLALAAVGL
                     AVMVADEIVKAATGVSFIQQALNPIMEHVLKPLMELIGKAITKALEGLGVDKKTAEMA
                     GSIVGAIVAAIAMVAVIVVVAVVGKGAAAKLGNALSKMMGETIKKLVPNVLKQLAQNG
                     SKLFTQGMQRITSGLGNVGSKMGLQTNALSKELVGNTLNKVALGMEVTNTAAQSAGGV
                     AEGVFIKNASEALADFMLARFAMDQIQQWLKQSVEIFGENQKVTAELQKAMSSAVQQN
                     ADASRFILRQSRA"
     gene            898. .2127
                     /gene="sspC"
     CDS             898. .2127
                     /gene="sspC"
                     /note="secreted protein"
                     /codon_start=1
                     /transl_table=11
                     /function="invasion determinant"
                     /evidence=experimental
                     /product="SspC"
                     /protein_id="AAC43547.1"
                     /db_xref="PID:g1117813"
                     /db_xref="GI:1117813"
                     /translation="MLISNVGINPAAYLNNHSVENSSQTASQSVSAKDILNSIGISSS
                     KVSDLGLSPTLSAPAPGVLTQTPGTITSFLKASIQNTDMNQDLNALANNVTTKANEVV
                     QTQLREQQAEVGKFFDISGMSSSAVALLAAANTLMLTLNQADSKLSGKLSLVSFDAAK
                     TTASSMMREGMNALSGSISQSALQLGITGVGAKLEYKGLQNERGALKHNAAKIDKLTT
                     ESHSIKNVLNGQNSVKLGAEGVDSLKSLNMKKTGTDATKNLNDATLKSNAGTSATESL
                     GIKDSNKQISPEHQAILSKRLESVESDIRLEQNTMDMTRIDARKMQMTGDLIMKNSVT
                     VGGIAGASGQYAATQERSEQQISQVNNRVASTASDEARESSRKSTSLIQEMLKTMESI
                     NQSKASALAAIAGNIRA"
     gene            2198. .3229
                     /gene="sspD"
     CDS             2198. .3229
                     /gene="sspD"
                     /note="secreted protein"
                     /codon_start=1
                     /transl_table=11
                     /function="invasion determinant"
                     /evidence=experimental
                     /product="SspD"
                     /protein_id="AAC43548.1"
                     /db_xref="PID:g1117814"
                     /db_xref="GI:1117814"
                     /translation="MLNIQNYSASPHPGIVAERPQTPSASEHVETAVVPSTTEHRGTD
                     IISLSQAATKIHQAQQTLQSTPPISEENNDERTLARQQLTSSLNALAKSGVSLSAEQN
                     ENLRSAFSAPTSALFSASPMAQPRTTISDAEIWDMVSQNISAIGDSYLGVYENVVAVY
                     TDFYQAFSDILSKMGGWLLPGKDGNTVKLDVTSLKNDLNSLVNKYNQINSNTVLFPAQ
                     SGSGVKVATEAEARQWLSELNLPNSCLKSYGSGYVVTVDLTPLQKMVQDIDGLGAPGK
                     DSKLEMDNAKYQAWQSGFKAQEENMKTTLQTLTQKYSNANSLYDNLVKVLSSTISSSL
                     ETAKSFLQG"

     gene            3248. .3441
                     /gene="sspA"
     CDS             3248. .>3441
                     /gene="sspA"
                     /note="secreted protein"
                     /codon_start=1
                     /transl_table=11
                     /function="unknown"
                     /product="SspA"
                     /protein_id="AAC43549.1"
                     /db_xref="PID:g1117815"
                     /db_xref="GI:1117815"
                     /translation="MVTSVRTQPPVIMPGMQTEIKTQATNLAANLSAVRESATATLSG
                     EIKGPQLEDFPALIKQASLDA"
BASE COUNT      978 a    759 c    925 g    779 t
ORIGIN
```

Query Match 100.0%; Score 870; DB 2; Length 3441;
Best Local Similarity 100.0%; Pred. No. 4.1e-203;
Matches 870; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 CGCAAAGCCGAGGAAACGAACCGCATTATGGGATGTATCGGGAAAGTCCTCGGCGCGCTG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CGCAAAGCCGAGGAAACGAACCGCATTATGGGATGTATCGGGAAAGTCCTCGGCGCGCTG 60

Qy      61 CTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCT 120

Qy     121 GCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCG 180

Qy     181 TTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTG 240

Qy     241 ATTGGCAAGGCGATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 ATTGGCAAGGCGATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAG 300

Qy     301 ATGGCCGGCAGCATTGTTGGTGCGATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 ATGGCCGGCAGCATTGTTGGTGCGATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTG 360

Qy     361 GTGGTCGCAGTTGTCGGGAAAGGCGCGGCGGCGAAACTGGGTAACGCGCTGAGCAAAATG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GTGGTCGCAGTTGTCGGGAAAGGCGCGGCGGCGAAACTGGGTAACGCGCTGAGCAAAATG 420

Qy     421 ATGGGCGAAACGATTAAGAAGTTGGTGCCTAACGTGCTGAAACAGTTGGCGCAAAACGGC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 ATGGGCGAAACGATTAAGAAGTTGGTGCCTAACGTGCTGAAACAGTTGGCGCAAAACGGC 480

Qy     481 AGCAAACTCTTTACCCAGGGGATGCAACGTATTACTAGCGGTCTGGGTAATGTGGGTAGC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 AGCAAACTCTTTACCCAGGGGATGCAACGTATTACTAGCGGTCTGGGTAATGTGGGTAGC 540

Qy     541 AAGATGGGCCTGCAAACGAATGCCTTAAGTAAAGAGCTGGTAGGTAATACCCTAAATAAA 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 AAGATGGGCCTGCAAACGAATGCCTTAAGTAAAGAGCTGGTAGGTAATACCCTAAATAAA 600

Qy     601 GTGGCGTTGGGCATGGAAGTCACGAATACCGCAGCCCAGTCAGCCGGTGGTGTTGCCGAG 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 GTGGCGTTGGGCATGGAAGTCACGAATACCGCAGCCCAGTCAGCCGGTGGTGTTGCCGAG 660

Qy     661 GGCGTATTTATTAAAAATGCCAGCGAGGCGCTTGCTGATTTTATGCTCGCCCGTTTTGCC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 GGCGTATTTATTAAAAATGCCAGCGAGGCGCTTGCTGATTTTATGCTCGCCCGTTTTGCC 720

Qy     721 ATGGATCAGATTCAGCAGTGGCTTAAACAATCCGTAGAAATATTTGGTGAAAACCAGAAG 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 ATGGATCAGATTCAGCAGTGGCTTAAACAATCCGTAGAAATATTTGGTGAAAACCAGAAG 780
```

```
Qy    781 GTAACGGCGGAACTGCAAAAAGCCATGTCTTCTGCGGTACAGCAAAATGCGGATGCTTCG 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GTAACGGCGGAACTGCAAAAAGCCATGTCTTCTGCGGTACAGCAAAATGCGGATGCTTCG 840

Qy    841 CGTTTTATTCTGCGCCAGAGTCGCGCATAA 870
          |||||||||||||||||||||||||||||||
Db    841 CGTTTTATTCTGCGCCAGAGTCGCGCATAA 870


RESULT  3
SDU66877
LOCUS       SDU66877     2995 bp    DNA             BCT       29-AUG-1996
DEFINITION  Salmonella dublin SicA (sicA) and SipB (sipB) genes, complete cds.
ACCESSION   U66877
NID         g1513311
VERSION     U66877.1  GI:1513311
KEYWORDS    .
SOURCE      Salmonella dublin.
  ORGANISM  Salmonella dublin
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Salmonella.
REFERENCE   1  (bases 1 to 2995)
  AUTHORS   Mullan,P.B., Gautier,A.V., Wood,M.W., Edwards,M.H., Jones,B.V. and
            Galyov,E.E.
  TITLE     Sip invasins of Salmonella dublin are involved in macrophage
            cytotoxicity and mouse virulence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2995)
  AUTHORS   Mullan,P.B., Gautier,A.V., Wood,M.W., Edwards,M.H., Jones,B.V. and
            Galyov,E.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-AUG-1996) Institute for Animal Health, Compton,
            Newbury, Berkshire, RG20 7NN, UK
FEATURES             Location/Qualifiers
     source          1. .2995
                     /organism="Salmonella dublin"
                     /strain="2229"
                     /db_xref="taxon:593"
     gene            131. .628
                     /gene="sicA"
     CDS             131. .628
                     /gene="sicA"
                     /codon_start=1
                     /transl_table=11
                     /product="SicA"
                     /protein_id="AAB06794.1"
                     /db_xref="PID:g1513312"
                     /db_xref="GI:1513312"
                     /translation="MDYQNNVSEERVAEMIWDAVSEGATLKDVHGIPQDMMDGLYAHA
                     YEFYNQGRLDEAETFFRFLCIYDFYNPDYTMGLAAVCQLKKQFQKACDLYAVAFTLLK
                     NDYRPVFFTGQCQLLMRKAAKARQCFELVNERTEDESLRAKALVYLEALKTAETEQHS
                     EQEKE"
     gene            631. .2412
                     /gene="sipB"
     CDS             631. .2412
                     /gene="sipB"
                     /function="virulence-associated secreted protein"
                     /codon_start=1
                     /transl_table=11
                     /product="SipB"
                     /protein_id="AAB06795.1"
                     /db_xref="PID:g1513313"
                     /db_xref="GI:1513313"
                     /translation="MVNDASSISRSGYTQNPRLAEAAFEGVRKNTDFLKAADKAFKDV
                     VATKAGDLKAGTKSGESAINTVGLKPPTDAAREKLSSEGQLTLLLGKLMTLLGDVSLS
                     QLESRLAVWQAMIESQKEMGIQVSKEFQTALGEAQEATDLYEASIKKTDTAKSVYDAA
                     TKKLTQAQNKLQSLDPADPGYAQAEAAVEQAGKEATEAKEALDKATDATVKAGTDAKA
                     KAEKADNILTKFQGTANAASQNQVSQGEQDNLSNVARLTMLMAMFIEIVGKNTEESLQ
                     NDLALFNALQEGRQAEMEKKSAEFQEETRKAEETNRIMGCIGKVLGALLTIVSVVAAV
                     FTGGASLALAAVGLAVMVADEIVKAATGVSFIQQALNPIMEHVLKPLMELIGKAITKA
                     LEGLGVDKKTAEMAGSIVGAIVAAIAMVAVIVVVAVVGKGAAAKLGNALSKMMGETIK
                     KLVPNVLKQLAQNGSKLFTQGMQRITSGLGNVGSKMGLQTNALSKELVGNTLNKVALG
                     MEVTNTAAQSAGGVAEGVFIKNASEALADFMLARFAMDQIQQWLKQSVEIFGENQKVT
                     AELQKAMSSAVQQNADASRFILRQSRA"
BASE COUNT      851 a    614 c    847 g    683 t
ORIGIN


  Query Match             99.4%;  Score 865.2;  DB 1;  Length 2995;
  Best Local Similarity   99.7%;  Pred. No. 6e-202;
  Matches  867;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 CGCAAAGCCGAGGAAACGAACCGCATTATGGGATGTATCGGGAAAGTCCTCGGCGCGCTG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1543 CGCAAAGCCGAGGAAACGAACCGCATTATGGGATGTATCGGGAAAGTCCTCGGCGCGCTG 1602

Qy     61 CTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1603 CTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCT 1662

Qy    121 GCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1663 GCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGGGTGTCG 1722

Qy    181 TTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1723 TTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTG 1782

Qy    241 ATTGGCAAGGCGATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAG 300
          ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db   1783 ATTGGCAAGGCGATTACCAAAGCGCTGGAAGGGTTAGGCGTCGATAAGAAAACGGCAGAG 1842

Qy    301 ATGGCCGGCAGCATTGTTGGTGCGATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1843 ATGGCCGGCAGCATTGTTGGTGCGATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTG 1902

Qy    361 GTGGTCGCAGTTGTCGGGAAAGGCGCGGCGGCGAAACTGGGTAACGCGCTGAGCAAAATG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1903 GTGGTCGCAGTTGTCGGGAAAGGCGCGGCGGCGAAACTGGGTAACGCGCTGAGCAAAATG 1962

Qy    421 ATGGGCGAAACGATTAAGAAGTTGGTGCCTAACGTGCTGAAACAGTTGGCGCAAAACGGC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1963 ATGGGCGAAACGATTAAGAAGTTGGTGCCTAACGTGCTGAAACAGTTGGCGCAAAACGGC 2022

Qy    481 AGCAAACTCTTTACCCAGGGGATGCAACGTATTACTAGCGGTCTGGGTAATGTGGGTAGC 540
          |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db   2023 AGCAAACTCTTTACCCAGGGGATGCAACGTATTACTAGCGGTCTGGGTAATGTAGGTAGC 2082

Qy    541 AAGATGGGCCTGCAAACGAATGCCTTAAGTAAAGAGCTGGTAGGTAATACCCTAAATAAA 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2083 AAGATGGGCCTGCAAACGAATGCCTTAAGTAAAGAGCTGGTAGGTAATACCCTAAATAAA 2142

Qy    601 GTGGCGTTGGGCATGGAAGTCACGAATACCGCAGCCCAGTCAGCCGGTGGTGTTGCCGAG 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2143 GTGGCGTTGGGCATGGAAGTCACGAATACCGCAGCCCAGTCAGCCGGTGGTGTTGCCGAG 2202

Qy    661 GGCGTATTTATTAAAAATGCCAGCGAGGCGCTTGCTGATTTTATGCTCGCCCGTTTTGCC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2203 GGCGTATTTATTAAAAATGCCAGCGAGGCGCTTGCTGATTTTATGCTCGCCCGTTTTGCC 2262

Qy    721 ATGGATCAGATTCAGCAGTGGCTTAAACAATCCGTAGAAATATTTGGTGAAAACCAGAAG 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2263 ATGGATCAGATTCAGCAGTGGCTTAAACAATCCGTAGAAATATTTGGTGAAAACCAGAAG 2322

Qy    781 GTAACGGCGGAACTGCAAAAAGCCATGTCTTCTGCGGTACAGCAAAATGCGGATGCTTCG 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2323 GTAACGGCGGAACTGCAAAAAGCCATGTCTTCTGCGGTACAGCAAAATGCGGATGCTTCG 2382

Qy    841 CGTTTTATTCTGCGCCAGAGTCGCGCATAA 870
          |||||||||||||||||||||||||||||||
Db   2383 CGTTTTATTCTGCGCCAGAGTCGCGCATAA 2412


RESULT  4
STSIPGNS
```

```
LOCUS       STSIPGNS     5393 bp    DNA             BCT       04-OCT-1995
DEFINITION  S.typhi spaT, sipB, sipC, sipD and sipA genes.
ACCESSION   X82670
NID         g1009001
VERSION     X82670.1  GI:1009001
KEYWORDS    sipA gene; sipB gene; sipC gene; sipD gene; spaT gene.
SOURCE      Salmonella typhi.
  ORGANISM  Salmonella typhi
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Salmonella.
REFERENCE   1  (bases 1 to 5393)
  AUTHORS   Hermant,D., Menard,R., Arricau,N., Parsot,C. and Popoff,M.Y.
  TITLE     Functional conservation of the Salmonella and Shigella effectors of
            entry into epithelial cells
  JOURNAL   Mol. Microbiol. 17 (4), 781-789 (1995)
  MEDLINE   96111497
REFERENCE   2  (bases 1 to 5393)
  AUTHORS   Popoff,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-1994) M. Popoff, Institut Pasteur, Unite des
            Enterobacteries, 28 rue du Docteur Roux, 75724 Paris Cedex 15,
            FRANCE
FEATURES             Location/Qualifiers
     source          1. .5393
                     /organism="Salmonella typhi"
                     /strain="Ty2"
                     /db_xref="taxon:601"
     gene            43. .540
                     /gene="spaT"
     CDS             43. .540
                     /gene="spaT"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA57987.1"
                     /db_xref="PID:g1009002"
                     /db_xref="GI:1009002"
                     /db_xref="SPTREMBL:Q57260"
                     /translation="MDYQNNVSEERVAEMIWDAVSEGATLKDVHGIPQDMMDGLYAHA
                     YEFYNQGRLDEAETFFRFLCIYDFYNPDYTMGLAAVCQLKKQFQKACDLYAVAFTLLK
                     NDYRPVFFTGQCQLLMRKAAKARQCFELVNERTEDESLRAKALVYLEALKTAETEQHS
                     EQEKE"
     gene            543. .2324
                     /gene="sipB"
     CDS             543. .2324
                     /gene="sipB"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA57988.1"
                     /db_xref="PID:g1009003"
                     /db_xref="GI:1009003"
                     /db_xref="SPTREMBL:Q56134"
                     /translation="MVNDASSISRSGYTQNPRLAEAAFEGVRKNTDFLKAADKAFKDV
                     VATKAGDLKAGTKSGESAINTVGLKPPTDAAREKLSSEGQLTLLLGKLMTLLGDVSLS
                     QLESRLAVWQAMIESQKEMGIQVSKEFQTALGEAQEATDLYEASIKKTDTAKSVYDAA
                     AKKLTQAQNKLQSLDPADPGYAQAEAAVEQAGKEATEAKEALDKATDATVKAGTDAKA
                     KAEKADNILTKFQGTANAASQNQVSQGEQDNLSNVARLTMLMAMFIEIVGKNTEESLQ
                     NDLALFNALQEGRQAEMEKKSAEFQEETRKAEETNRIMGCIGKVLGALLTIVSVVAAV
                     FTGGASLALAAVGLAVMVADEIVKAATGVSFIQQALNPIMEHVLKPLMELIGKAITKA
                     LEGLGVDKKTAEMAGSIVGAIVAAIAMVAVIVVVAVVGKGAAAKLGNALSKMMGETIK
                     KLVPNVLKQLAQNGSKLFTQGMQRITSGLGNVGSKMGLQTNALSKELVGNTLNKVALG
                     MEVTNTAAQSAGGVAEGVFIKNASEALADFMLARFAMDQIQQWLKQSVEIFGENQKVT
                     AELQKAMSSAVQQNADASRFILRQSRA"
     gene            2352. .3581
                     /gene="sipC"
     CDS             2352. .3581
                     /gene="sipC"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA57989.1"
                     /db_xref="PID:g1009004"
                     /db_xref="GI:1009004"
                     /db_xref="SPTREMBL:Q56135"
                     /translation="MLISNVGINPAAYLNNHSVENSSQTASQSVSAKDILNSIGISSS
                     KVSDLGLSPTLSAPAPGVLTQTPGTITSFLKASIQNTDMNQDLNALANNVTTKANEVV
                     QTQLREQQAEVGKFFDISGMSSSAVALLAAANILMLTLNQADSKLSGKLSLVSFDAAK
                     TTASSMMREGMNALSGSISQSALQLGITGVGAKLEYKGLQNERGALKHNAAKIDKLTT
                     ESHSIKNVLNGQNSVKLGAEGVDSLKSLNMKKTGTDATKNLNDATLKSNAGTSATESL
                     GIKNSNKQISPEHQAILSKRLESVESDIRLEQNTMDMTRIDARKMQMTGDLIMKNSVT
                     VGGIAGASRQYAATQERSEQQISQVNNRVASTASDEARESSRKSTSLIQEMLKTMESI
                     NQSKASALAAIAGNIRA"
     gene            3651. .4670
                     /gene="sipD"
     CDS             3651. .4670
                     /gene="sipD"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA57990.1"
                     /db_xref="PID:g1009005"
                     /db_xref="GI:1009005"
                     /db_xref="SPTREMBL:Q56136"
                     /translation="MLNIQNYSASPHPGIVAERPQTPSASEHAEIAVVPSTTEHRGTD
                     IISLSQAATKIQQAQQTLQSTPPISEENNDERTLARQQLTSSLNALAKSGVSLSAEQN
                     ENLRSTFSARRRPYLALRLWPARTTISDAEIWDMVSQNISAIGDSYLGVYENVVAVYT
                     DFYQAFSDILSKMGGWLSPGKDGNTIKLNVDSLKSEISSLINKYTQINKNTILFPSQT
                     GSGMTTATKAEAEQWIKELNLPDSCLKASGSGYVVLVDTGPLSKMVSDLNGIGSGSAL
                     ELDNAKYQAWQSGFKAQEENLKTTLQTLTQKYSNANSLYDNLVKVLSSTISSSLETAK
                     SFLQG"
     gene            4689. .5393
                     /gene="sipA"
     CDS             4689. .>5393
                     /gene="sipA"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA57991.1"
                     /db_xref="PID:g1009006"
                     /db_xref="GI:1009006"
                     /db_xref="SPTREMBL:Q56137"
                     /translation="MVTSVRTQPPVIMPGMQTEIKTQATNLAANLSAVRESATATLSG
                     EIKGQQLEDFPALIKQASLDALFKCGKDAEALKEVFTNSNNVAGKKAIMEFAGLFRSA
                     LNATSDSPEAKTLLMKVGAEYTAQIIKDGLKEKSAFGPWLPETKKAEAKLENLEKQLL
                     DIIKNNTGGELSKLSTNLVMQEVMPYIASCIEHNFGCTLDPLTRSSLTQLVDKAAAKA
                     VEALDMCHQKLTQEQGT"
BASE COUNT     1584 a   1156 c   1458 g   1195 t
ORIGIN
```

```
  Query Match             98.7%;  Score 858.8;  DB 1;  Length 5393;
  Best Local Similarity   99.2%;  Pred. No. 2.4e-200;
  Matches  863;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      1 CGCAAAGCCGAGGAAACGAACCGCATTATGGGATGTATCGGGAAAGTCCTCGGCGCGCTG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1455 CGCAAAGCCGAGGAAACGAACCGCATTATGGGATGTATCGGGAAAGTCCTCGGCGCGCTG 1514

Qy     61 CTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1515 CTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCT 1574

Qy    121 GCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1575 GCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGGGTGTCG 1634

Qy    181 TTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1635 TTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTG 1694

Qy    241 ATTGGCAAGGCGATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1695 ATTGGCAAGGCGATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAG 1754

Qy    301 ATGGCCGGCAGCATTGTTGGTGCGATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTG 360
          ||||| ||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db   1755 ATGGCAGGCAGCATTGTTGGTGCGATTGTCGCCGCTATTGCCATGGTAGCGGTCATTGTG 1814

Qy    361 GTGGTCGCAGTTGTCGGGAAAGGCGCGGCGGCGAAACTGGGTAACGCGCTGAGCAAAATG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   1815 GTGGTCGCAGTTGTCGGGAAAGGCGCGGCGGCGAAACTGGGTAACGCGCTGAGCAAAATG 1874

Qy    421 ATGGGCGAAACGATTAAGAAGTTGGTGCCTAACGTGCTGAAACAGTTGGCGCAAAACGGC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db   1875 ATGGGCGAAACGATTAAGAAGTTGGTGCCTAACGTGCTGAAACAGTTGGCACAAAACGGC 1934

Qy    481 AGCAAACTCTTTACCCAGGGGATGCAACGTATTACTAGCGGTCTGGGTAATGTGGGTAGC 540
          |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db   1935 AGCAAACTCTTTACCCAGGGGATGCAACGTATTACTAGCGGCCTGGGTAATGTGGGTAGC 1994

Qy    541 AAGATGGGCCTGCAAACGAATGCCTTAAGTAAAGAGCTGGTAGGTAATACCCTAAATAAA 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1995 AAGATGGGCCTGCAAACGAATGCCTTAAGTAAAGAGCTGGTAGGTAATACCCTAAATAAA 2054

Qy    601 GTGGCGTTGGGCATGGAAGTCACGAATACCGCAGCCCAGTCAGCCGGTGGTGTTGCCGAG 660
          |||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db   2055 GTGGCGTTGGGCATGGAAGTCACGAATACCGCAGCCCAGTCAGCCGGTGGGGTTGCCGAG 2114

Qy    661 GGCGTATTTATTAAAAATGCCAGCGAGGCGCTTGCTGATTTTATGCTCGCCCGTTTTGCC 720
          || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2115 GGGGTATTTATTAAAAATGCCAGCGAGGCGCTTGCTGATTTTATGCTCGCCCGTTTTGCC 2174

Qy    721 ATGGATCAGATTCAGCAGTGGCTTAAACAATCCGTAGAAATATTTGGTGAAAACCAGAAG 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2175 ATGGATCAGATTCAGCAGTGGCTTAAACAATCCGTAGAAATATTTGGTGAAAACCAGAAG 2234

Qy    781 GTAACGGCGGAACTGCAAAAAGCCATGTCTTCTGCGGTACAGCAAAATGCGGATGCTTCG 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2235 GTAACGGCGGAACTGCAAAAAGCCATGTCTTCTGCGGTACAGCAAAATGCGGATGCTTCG 2294

Qy    841 CGTTTTATTCTGCGCCAGAGTCGCGCATAA 870
          |||||||||||||||||||||||||||||
Db   2295 CGTTTTATTCTGCGCCAGAGTCGCGCATAA 2324
```

```
RESULT   5
SDIPABCD
LOCUS       SDIPABCD     6869 bp    DNA             BCT       24-JUL-1991
DEFINITION  Shigella dysenteriae ipaBCD genes.
ACCESSION   X60777
NID         g46932
VERSION     X60777.1  GI:46932
KEYWORDS    ipaB gene; ipaC gene; ipaD gene; ippI gene.
SOURCE      Shigella dysenteriae.
  ORGANISM  Shigella dysenteriae
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Shigella.
REFERENCE   1  (bases 1 to 6869)
  AUTHORS   Palchaudhuri,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUN-1991) S. Palchaudhuri, Wayne State Univers.School
            of Medicine, Department of Immunology and, Microbiology, 540 E.
            Canfield, Detroit, MI 48201, USA
REFERENCE   2  (bases 1 to 6869)
  AUTHORS   Yao,R. and Palchaudhuri,S.
  TITLE     Nucleotide sequence of the ipaBCD structure genes of Shigella
            dysenteriae
  JOURNAL   Unpublished
FEATURES             Location/Qualifiers
     source          1. .6869
                     /organism="Shigella dysenteriae"
                     /strain="CG097"
                     /db_xref="taxon:622"
     CDS             317. .943
                     /note="24kD ORF"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA43188.1"
                     /db_xref="PID:g46933"
                     /db_xref="GI:46933"
                     /db_xref="SWISS-PROT:Q04314"
                     /translation="MQILNKILPQVEFAIPRPSFNSLSRNKLVKKILSVFNLKQRFPQ
                     KNFGCTVNINKIRDNVIDKIKDSNSGNQLFCWMSQERTTYVSSMINRSIDEMAIHNGV
                     VLTSDNKKNIFAAIEKKFPDIKLDEKSAQTSISHTALNEIASSGLRAKILKRYSSDMD
                     LFNTQMKDLTNLVSSSVYDKIFNESTKVLQIEISAEVLKAVYRQSNTN"
     CDS             1000. .1467
                     /note="IppI ORF"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA43189.1"
                     /db_xref="PID:g46934"
                     /db_xref="GI:46934"
                     /db_xref="SWISS-PROT:P18008"
                     /translation="MSLNITENESISTAVIDAINSGATLKDINAIPDDMMDDIYSYAY
                     DFYNKGRIEEAEVFFRFLCIYDFYNVDYIMGLAAIYQIKEQFQQAADLYAVAFALGKN
                     DYTPVFHTGQCQLRLKAPLKAKECFELVIQHSNDEKLKIKAQSYLDAIQDIKE"
     CDS             1473. .3215
                     /note="IpaB ORF"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA43190.1"
                     /db_xref="PID:g46935"
                     /db_xref="GI:46935"
                     /db_xref="SWISS-PROT:Q03945"
                     /translation="MHNVNTTTTGLSLAKILASTELGDNTIQAGNDAANKLFSLTIAD
                     LTANKNINTTNAHSTSNILIPELKAPKSLNASSQLTLLIGNLIQILGEKSLTALTNKI
                     TAWKSQQQARQQKNLEFSDKINTLLSETEGLIRDYEKQINKLKNADSKIKDLENKINQ
                     IQTRLSELDPDSPEKKKLSREEIQLTIKKDAAVKDRTLIEQKTLSIHSKLTDKSMQLE
                     KEIDSFSAFSNTASAEQLSTQQKSLTGLASVTQLMATFIQLVGKNNEESLKNDLALFQ
                     SLQESRKTEMERKSDEYAAEVRKAEELNRVMGCVGKILGALLTIVSVVAAAFSGGASL
                     ALADVGLALMVTDAIVQAATGNSFMEQALNPIMKAVIEPLIKLLSDAFTKMLEGLGVD
                     SKKAKMIGSILGAIAGALVLVAAVVLVATVGKQAAAKLAENIGKIIGKTLTDLIPKFL
                     KNFSSQLDDLITNAVARLNKFLGAAGDEVISKQIISTHLNQAVLLGESVNSATQAGGS
                     VASAVFQNSASTNLADLILSKYQVEQLSKYISEAIEKFGQLQEVIADLLASMSNSQAN
                     RTDVAKAILQQTTA"
     CDS             3178. .4326
                     /note="IpaC ORF"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA43191.1"
                     /db_xref="PID:g46936"
                     /db_xref="GI:46936"
                     /db_xref="SWISS-PROT:Q03946"
                     /translation="MLQKQFCNKLLLDTNKENVMEIQNTKSTQILYTDISTKQTQSSS
                     ETQKSQNYQQIAAHIPLNVGKNPVLTTTLNDDQLLKLSEQVQHDSEIIARLTDKKMKD
                     LSEMSHTLTPENTLDISSLSSNAVSLIISVAVLLSALRTAETKLGSQLSLIAFDATKS
                     AAENIVRQGLAALSSSITGAVTQVGITGIGAKKTHSGISDQKGALRKNLATAQSLEKE
                     LAGSKLGLNKQIDTNITSPQTNSSTKFLGKNKLAPDNISLSTEHKTSLSSPDISLQDK
                     IDTQRRTYELNTLSAQQKQNIGRATMETSAVAGNISTSGGRYASALEEEEQLISQASS
                     KQAEEASQVSKEASQATNQLIQKLLNIIDNINQSRSSTASQIAGNIRA"
     CDS             4377. .5375
                     /note="IpaD ORF"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA43192.1"
                     /db_xref="PID:g46937"
                     /db_xref="GI:46937"
                     /db_xref="SWISS-PROT:Q03947"
                     /translation="MNITTLTNSISTSSFSPNNTNGSSTETVNSDIKTTTSSHPVSSL
                     TMLNDTLHNIRTTNQALKKELSQKTLTKTSLEEIALHSSQISMDVNKSAQLLNILSKT
                     EYPINKDARELLHSAPKEAELDGYEMISHRELWAKIANSINDINEQYLKVYEHAVSSY
                     TQMYQEFSAVLSSLAGWISPGGNDGNSVKLQVKSLKDALTTLKKNYEDKPLYPATNTV
                     SEQEANKWLTELGGTIGTVSAKNGGYVVSINMTPIYNMLNRLDNLGGNGEVVLDNAKY
                     QAWNAGFSAEDETMKNNLQTLVQKYSNANSIFDNLVKVLSSTISSCTDTDKLFLHF"
BASE COUNT     2518 a   1369 c   1135 g   1847 t
ORIGIN
```

```
  Query Match             20.5%;  Score 178.6;  DB 1;  Length 6869;
  Best Local Similarity   58.3%;  Pred. No. 7e-34;
  Matches  313;  Conservative    0;  Mismatches  224;  Indels    0;  Gaps    0;

Qy       1 CGCAAAGCCGAGGAAACGAACCGCATTATGGGATGTATCGGGAAAGTCCTCGGCGCGCTG 60
           || ||||| || |||   ||| |  | ||||| ||| | |||||| | || || || ||
Db    2364 CGTAAAGCAGAAGAACTCAACAGAGTAATGGGTTGTGTTGGGAAAATACTTGGGGCACTT 2423
```

```
Qy      61 CTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCT 120
Db    2424 TTAACTATCGTTAGTGTTGTTGCAGCAGCTTTTTCTGGAGGAGCCTCTCTAGCACTGGCA 2483

Qy     121 GCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCG 180
Db    2484 GATGTTGGTTTAGCTCTTATGGTTACGGATGCTATAGTACAAGCAGCGACCGGCAATTCC 2543

Qy     181 TTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTG 240
Db    2544 TTCATGGAACAAGCCCTGAATCCGATCATGAAAGCAGTCATTGAACCCTTAATCAAACTC 2603

Qy     241 ATTGGCAAGGCGATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAG 300
Db    2604 CTTTCAGATGCATTTACAAAAATGCTCGAAGGCTTGGGCGTCGACTCGAAAAAAGCCAAA 2663

Qy     301 ATGGCCGGCAGCATTGTTGGTGCGATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTG 360
Db    2664 ATGATTGGCTCTATTCTGGGGGCAATCGCAGGCGCTCTTGTCCTGGTTGCAGCAGTCGTT 2723

Qy     361 GTGGTCGCAGTTGTCGGGAAAGGCGCGGCGGCGAAACTGGGTAACGCGCTGAGCAAAATG 420
Db    2724 CTCGTAGCCACTGTTGGTAAACAGGCAGCAGCAAAACTTGCAGAGAATATTGGCAAAATA 2783

Qy     421 ATGGGCGAAACGATTAAGAAGTTGGTGCCTAACGTGCTGAAACAGTTGGCGCAAAACGGC 480
Db    2784 ATAGGTAAAACCCTCACAGACCTTATACCAAAGTTTCTCAAGAATTTTTCTTCTCAACTG 2843

Qy     481 AGCAAACTCTTTACCCAGGGGATGCAACGTATTACTAGCGGTCTGGGTAATGTGGGT 537
Db    2844 GACGATTTAATCACTAATGCTGTTGCCAGATTAAATAAATTTCTTGGTGCAGCGGGT 2900
```

RESULT 6
SFPVR2
LOCUS SFPVR2 7663 bp DNA BCT 12-SEP-1993
DEFINITION Shigella flexineri plasmid YSH6000 virulence region-2 DNA containing 24k gene, 18k gene, ipaB gene, ipaC gene and ipaD gene.
ACCESSION X15319
NID g47052
VERSION X15319.1 GI:47052
KEYWORDS ipaB gene; ipaC gene; ipaD gene; virulence proteins.
SOURCE Enterobacteriaceae.
ORGANISM Enterobacteriaceae
Bacteria; Proteobacteria; gamma subdivision.
REFERENCE 1 (bases 1 to 7663)
AUTHORS Sasakawa,C.
TITLE Direct Submission
JOURNAL Submitted (19-MAY-1989) Sasakawa C., Institute of Medical Science, University of Tokyo, Shirokanedai 4-6-1, Minato-ku Tokyo 108, Japan
REFERENCE 2 (bases 1010 to 6538)
AUTHORS Sasakawa,C., Adler,B., Tobe,T., Okada,N., Nagai,S., Komatsu,K. and Yoshikawa,M.
TITLE Functional organization and nucleotide sequence of virulence Region-2 on the large virulence plasmid in Shigella flexneri 2a
JOURNAL Mol. Microbiol. 3 (9), 1191-1201 (1989)
MEDLINE 90014179
COMMENT Region 2 has three promoters that transcribe RNAs of 4.0, 4.5 and 7.5kb. The 4.0 kb transcript produces the 24,18kD, IpaB and IpaC proteins. The 4.5 kb transcript produces the ipaD protein. The 7.5kb transcript covers the entire region 2.
Data kindly reviewed (20-NOV-1989) by Sasakawa C.

FEATURES Location/Qualifiers

```
     source          1. .7663
                     /organism="Enterobacteriaceae"
                     /plasmid="YSH6000"
                     /db_xref="taxon:543"
     promoter        1291. .1296
                     /note="-35 region for 4kb transcript"
     promoter        1312. .1318
                     /note="-10 region for 4kb transcript"
     RBS             1360. .1362
                     /note="put. ribosome binding site"
     CDS             1370. .1996
                     /note="24k gene (AA 1-208)"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA33379.1"
                     /db_xref="PID:g47053"
                     /db_xref="GI:47053"
                     /db_xref="SWISS-PROT:P33548"
                     /translation="MQILNKILPQVEFAIPRPSFDSLSRNKLVKKILSVFNLKQRFPQ
                     KNFGCPVNINKIRDSVIDKIKDSNSGNQLFCWMSQERTTYVSSMINRSIDEMAIHNGV
                     VLTSDNKKNIFAAIEKKFPDIKLDEKSAQTSISHTALNEIASSGLRAKILKRYSSDMD
                     LFNTQMKDLTNLVSSSVYDKIFNESTKVLQIEISAEVLKAVYRQSNTN"
     RBS             2043. .2045
                     /note="put. ribosome binding site"
     CDS             2053. .2520
                     /note="18k gene (AA 1-155)"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA33380.1"
                     /db_xref="PID:g47054"
                     /db_xref="GI:47054"
                     /db_xref="SWISS-PROT:P18008"
                     /translation="MSLNITENESISTAVIDAINSGATLKDINAIPDDMMDDIYSYAY
                     DFYNKGRIEEAEVFFRFLCIYDFYNVDYIMGLAAIYQIKEQFQQAADLYAVAFALGKN
                     DYTPVFHTGQCQLRLKAPLKAKECFELVIQHSNDEKLKIKAQSYLDAIQDIKE"
     RBS             2513. .2515
                     /note="put. ribosome binding site"
     CDS             2526. .4268
                     /note="ipaB gene (AA 1-580)"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA33381.1"
                     /db_xref="PID:g47055"
                     /db_xref="GI:47055"
                     /db_xref="SWISS-PROT:P18011"
                     /translation="MHNVSTTTGFPLAKILASTELGDNTIQAANDAANKLFSLTIAD
                     LTANQNINTTNAHSTSNILIPELKAPKSLNASSQLTLLIGNLIQILGEKSLTALTNKI
                     TAWKSQQQARQQKNLEFSDKINTLLSETEGLTRDYEKQINKLKNADSKIKDLENKINQ
                     IQTRLSELDPESPEKKKLSREEIQLTIKKDAAVKDRTLIEQKTLSIHSKLTDKSMQLE
                     KEIDSFSAFSNTASAEQLSTQQKSLTGLASVTQLMATFIQLVGKNNEESLKNDLALFQ
                     SLQESRKTEMERKSDEYAAEVRKAEELNRVMGCVGKILGALLTIVSVVAAAFSGGASL
                     ALAAVGLALMVTDAIVQAATGNSFMEQALNPIMKAVIEPLIKLLSDAFTKMLEGLGVD
                     SKKAKMIGSILGAIAGALVLVAAVVLVATVGKQAAAKLAENIGKIIGKTLTDLIPKFL
                     KNFSSQLDDLITNAVARLNKFLGAAGDEVISKQIISTHLNQAVLLGESVNSATQAGGS
                     VASAVFQNSASTNLADLTLSKYQVEQLSKYISEAIEKFGQLQEVIADLLASMSNSQAN
                     RTDVAKAILQQTTA"
     RBS             4224. .4226
                     /note="put. ribosome binding site"
     CDS             4231. .5379
                     /note="ipaC gene (AA 1-382)"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA33382.1"
                     /db_xref="PID:g47056"
                     /db_xref="GI:47056"
                     /db_xref="SWISS-PROT:P18012"
                     /translation="MLQKQFCNKLLLDTNKENVMEIQNTKPTQILYTDISTKQTQSSS
                     ETQKSQNYQQIAAHIPLNVGKNPVLTTTLNDDQLLKLSEQVQHDSEIIARLTDKKMKD
                     LSEMSHTLTPENTLDISSLSSNAVSLIISVAVLLSALRTAETKLGSQLSLIAFDATKS
                     AAENIVRQGLAALSSSITGAVTQVGITGIGAKKTHSGISDQKGALRKNLATAQSLEKE
                     LAGSKLGLNKQIDTNITSPQTNSSTKFLGKNKLAPDNISLSTEHKTSLSSPDISLQDK
                     IDTQRRTYELNTLSAQQKQNIGRATMETSAVAGNISTSGGRYASALEEEEQLISQASS
                     KQAEEASQVSKEASQATNQLIQKLLNIIDSINQSKNSTASQIAGNIRA"
     promoter        5376. .5381
                     /note="-35 region for 4.5kb transcript"
     promoter        5396. .5401
                     /note="-10 region for 4.5kb transcript"
     RBS             5421. .5423
                     /note="put. ribosome binding site"
     CDS             5430. .6428
                     /note="ipaD gene (AA 1-332)"
                     /codon_start=1
                     /transl_table=11
```

```
                     /protein_id="CAA33383.1"
                     /db_xref="PID:g47057"
                     /db_xref="GI:47057"
                     /db_xref="SWISS-PROT:P18013"
                     /translation="MNITTLTNSISTSSFSPNNTNGSSTETVNSDIKTTTSSHPVSSL
                     TMLNDTLHNIRTTNQALKKELSQKTLTKTSLEEIALHSSQISMDVNKSAQLLDILSRH
                     EYPINKDARELLHSAPKEAELDGDQMISHRELWAKIANSINDINEQYLKVYEHAVSSY
                     TQMYQDFSAVLSSLAGWISPGGNDGNSVKLQVNSLKKALEELKEKYKDKPLYPANNTV
                     SQEQANKWLTELGGTIGKVSQKNGGYVVSINMTPIDNMLKSLDNLGGNGEVVLDNAKY
                     QAWNAGFSAEDETMKNNLQTLVQKYSNANSIFDNLVKVLSSTISSCTDTDKLFLHF"
BASE COUNT     2845 a   1501 c   1246 g   2071 t
ORIGIN


  Query Match           20.5%;  Score 178.6;  DB 1;  Length 7663;
  Best Local Similarity  58.3%;  Pred. No. 7.1e-34;
  Matches  313;  Conservative    0;  Mismatches  224;  Indels    0;  Gaps    0;

Qy       1 CGCAAAGCCGAGGAAACGAACCGCATTATGGGATGTATCGGGAAAGTCCTCGGCGCGCTG 60
             || ||||| || |||   ||| |  | ||||| ||| | |||||| | || || || ||
Db    3417 CGTAAAGCAGAAGAACTCAACAGAGTAATGGGTTGTGTTGGGAAAATACTTGGGGCACTT 3476

Qy      61 CTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCT 120
            |||| || || || ||||| || || | |||| | || || ||   ||| || |||||
Db    3477 TTAACTATCGTTAGTGTTGTTGCAGCAGCTTTTTCTGGAGGAGCCTCTCTAGCACTGGCA 3536

Qy     121 GCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCG 180
           || || ||  | ||  | |||||  | ||||  || ||  | || ||||| ||     ||
Db    3537 GCTGTTGGTTTAGCTCTTATGGTTACGGATGCTATAGTACAAGCAGCGACCGGCAATTCC 3596

Qy     181 TTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTG 240
           || ||  | || || || || ||||| ||| |     ||  |  | || |||||  | ||
Db    3597 TTCATGGAACAAGCCCTGAATCCGATCATGAAAGCAGTCATTGAACCCTTAATCAAACTC 3656

Qy     241 ATTGGCAAGGCGATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAG 300
             ||    | ||  |||| |||  ||| ||||| || ||||||||   |||||  ||  |
Db    3657 CTTTCAGATGCATTTACAAAAATGCTCGAAGGCTTGGGCGTCGACTCGAAAAAAGCCAAA 3716

Qy     301 ATGGCCGGCAGCATTGTTGGTGCGATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTG 360
           |||   |||   ||| | || || || |  | |||| ||| | | || || |   | ||
Db    3717 ATGATTGGCTCTATTCTGGGGGCAATCGCAGGCGCTCTTGTCCTAGTTGCAGCAGTCGTT 3776

Qy     361 GTGGTCGCAGTTGTCGGGAAAGGCGCGGCGGCGAAACTGGGTAACGCGCTGAGCAAAATG 420
             | || ||   ||| || |||   || || || ||||| |   |     |  |||||||
Db    3777 CTCGTAGCCACTGTTGGTAAACAGGCAGCAGCAAAACTTGCAGAAAATATTGGCAAAATA 3836

Qy     421 ATGGGCGAAACGATTAAGAAGTTGGTGCCTAACGTGCTGAAACAGTTGGCGCAAAACGGC 480
           || ||  ||||  | |   |  |  | || ||  | || ||  | ||  |      |
Db    3837 ATAGGTAAAACCCTCACAGACCTTATACCAAAGTTTCTCAAGAATTTTTCTTCTCAACTG 3896

Qy     481 AGCAAACTCTTTACCCAGGGGATGCAACGTATTACTAGCGGTCTGGGTAATGTGGGT 537
             | |  |  | ||  | |   |     |  | | ||    ||| |||   | ||||
Db    3897 GACGATTTAATCACTAATGCTGTTGCCAGATTAAATAAATTTCTTGGTGCAGCGGGT 3953


RESULT   7
SHFIPABCD
LOCUS       SHFIPABCD    5255 bp    DNA             BCT       19-MAR-1991
DEFINITION  S.flexneri ipaBCD region encoding invasion plasmid antigens.
ACCESSION   J04117
NID         g152740
VERSION     J04117.1  GI:152740
KEYWORDS    ipaA gene; ipaB gene; ipaC gene; ipaD; membrane protein.
SOURCE      S.flexeri (strain M90T-W) DNA, clone pHC17.
  ORGANISM  Shigella flexneri
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Shigella.
REFERENCE   1  (bases 1 to 5255)
  AUTHORS   Venkatesan,M.M., Buysse,J.M. and Kopecko,D.J.
  TITLE     Characterization of invasion plasmid antigen genes (ipaBCD) from
            Shigella flexneri
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 85, 9317-9321 (1988)
  MEDLINE   89057927

COMMENT     Draft entry and clean copy of sequence [1] kindly submitted by
            M.M.Venkatesan, 03-JAN-1989.
FEATURES             Location/Qualifiers
     source          1. .5255
                     /organism="Shigella flexneri"
                     /db_xref="taxon:623"
     CDS             606. .1073
                     /note="18 kD protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA26521.1"
                     /db_xref="PID:g152741"
                     /db_xref="GI:152741"
                     /translation="MSLNITENESISTAVIDAINSGATLKDINAIPDDMMDDIYSYAY
                     DFYNKGRIEEAEVFFRFLCIYDFYNVDYIMGLAAIYQIKEQFQQAADLYAVAFALGKN
                     DYTPVFHTGQCQLRLKAPLKAKECFELVIQHSNDEKLKIKAQSYLDAIQDIKE"
     CDS             1079. .2821
                     /note="IpaB protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA26522.1"
                     /db_xref="PID:g152742"
                     /db_xref="GI:152742"
                     /translation="MHNVSTTTGFPLAKILTSTELGDNTIQAANDAANKLFSLTIAD
                     LTANQNINTTNAHSTSNILIPELKAPKSLNASSQLTLLIGNLIQILGEKSLTALTNKI
                     TAWKSQQQARQQKNLEFSDKINTLLSETEGLTRDYEKQINKLKNADSKIKDLENKINQ
                     IQTRLSNLDPESPEKKKLSREEIQLTIKKDAAVKDRTLIEQKTLSIHSKLTDKSMQLE
                     KEIDSFSAFSNTASAEQLSTQQKSLTGLASVTQLMATFIQLVGKNNEESLKNDLALFQ
                     SLQESRKTEMERKSDEYAAEVRKAEELNRVMGCVGKILGALLTIVSVVAAAFSGGASL
                     ALAAVGLALMVTDAIVQAATGNSFMEQALNPIMKAVIEPLIKLLSDAFTKMLEGLGVD
                     SKKAKMIGSILGAIAGALVLVAAVVLVATVGKQAAAKLAENIGKIIGKTLTDLIPKFL
                     KNFSSQLDDLITNAVARLNKFLGAAGDEVISKQIISTHLNQAVLLGESVNSATQAGGS
                     VASAVFQNSASTNLADLTLSKYQVEQLSKYISEAIEKFGQLQEVIADLLASMSNSQAN
                     RTDVAKAILQQTTA"
     CDS             2784. .3932
                     /note="IpaC protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA26523.1"
                     /db_xref="PID:g152743"
                     /db_xref="GI:152743"
                     /translation="MLQKQFCNKLLLDTNKENVMEIQNTKPTQTLYTDISTKQTQSSS
                     ETQKSQNYQQIAAHIPLNVGKNPVLTTTLNDDQLLKLSEQVQHDSEIIARLTDKKMKD
                     LSEMSHTLTPENTLDISSLSSNAVSLIISVAVLLSALRTAETKLGSQLSLIAFDATKS
                     AAENIVRQGLAALSSSITGAVTQVGITGIGAKKTHSGISDQKGALRKNLATAQSLEKE
                     LAGSKLGLNKQIDTNITSPQTNSSTKFLGKNKLAPDNISLSTEHKTSLSSPDISLQDK
                     IDTQRRTYELNTLSAQQKQNIGRATMETSAVAGNISTSGGRYASALEEEEQLISQASS
                     KQAEEASQVSKEASQATNQLIQKLLNIIDSINQSKNSAASQIAGNIRA"
     CDS             3983. .4981
                     /note="IpaD protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA26524.1"
                     /db_xref="PID:g152744"
                     /db_xref="GI:152744"
                     /translation="MNITTLTNSISTSSFSPNNTNGSSTETVNSDIKTTTSSHPVSSL
                     TMLNDTLHNIRTTNQALKKELSQKTLTKTSLEEIALHSSQISMDVNKSAQLLDILSRN
                     EYPINKDARELLHSAPKEAELDGDQMISHRELWAKIANSINDINEQYLKVYEHAVSSY
                     TQMYQDFSAVLSSLAGWISPGGNDGNSVKLQVNSLKKALEELKEKYKDKPLYPANNTV
                     SQEQANKWLTELGGTIGKVSQKNGGYVVSINMTPIDNMLKSLDNLGGNGEVVLDNAKY
                     QAWNAGFSAEDETMKNNLQTLVQKYSNANSIFDNLVKVLSSTISSCTDTDKLFLHF"
     CDS             4990. .>5255
                     /note="IpaA protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA26525.1"
                     /db_xref="PID:g551981"
                     /db_xref="GI:551981"
                     /translation="MHNVNNTQAPIFLYKATSPSSTEYSELKSKISDIHSSQTSLKTP
                     ASVSEKENFATSFNQKCLDFLFSSSGKEDVLRSIYSNSMNAYAK"
BASE COUNT     1950 a   1037 c    889 g   1379 t
ORIGIN
```

```
  Query Match             20.5%;  Score 178.6;  DB 1;  Length 5255;
  Best Local Similarity   58.3%;  Pred. No. 6.7e-34;
  Matches  313;  Conservative    0;  Mismatches  224;  Indels    0;  Gaps    0;

Qy       1 CGCAAAGCCGAGGAAACGAACCGCATTATGGGATGTATCGGGAAAGTCCTCGGCGCGCTG 60
Db    1970 CGTAAAGCAGAAGAACTCAACAGAGTAATGGGTTGTGTTGGGAAAATACTTGGGGCACTT 2029

Qy      61 CTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCT 120
Db    2030 TTAACTATCGTTAGTGTTGTTGCAGCAGCTTTTTCTGGAGGAGCCTCTCTAGCACTGGCA 2089

Qy     121 GCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCG 180
Db    2090 GCTGTTGGTTTAGCTCTTATGGTTACGGATGCTATAGTACAAGCAGCGACCGGCAATTCC 2149

Qy     181 TTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTG 240
Db    2150 TTCATGGAACAAGCCCTGAATCCGATCATGAAAGCAGTCATTGAACCCTTAATCAAACTC 2209

Qy     241 ATTGGCAAGGCGATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAG 300
Db    2210 CTTTCAGATGCATTTACAAAAATGCTCGAAGGCTTGGGCGTCGACTCGAAAAAAGCCAAA 2269

Qy     301 ATGGCCGGCAGCATTGTTGGTGCGATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTG 360
Db    2270 ATGATTGGCTCTATTCTGGGGGCAATCGCAGGCGCTCTTGTCCTAGTTGCAGCAGTCGTT 2329

Qy     361 GTGGTCGCAGTTGTCGGGAAAGGCGCGGCGGCGAAACTGGGTAACGCGCTGAGCAAAATG 420
Db    2330 CTCGTAGCCACTGTTGGTAAACAGGCAGCAGCAAAACTTGCAGAAAATATTGGCAAAATA 2389

Qy     421 ATGGGCGAAACGATTAAGAAGTTGGTGCCTAACGTGCTGAAACAGTTGGCGCAAAACGGC 480
Db    2390 ATAGGTAAAACCCTCACAGACCTTATACCAAAGTTTCTCAAGAATTTTTCTTCTCAACTG 2449

Qy     481 AGCAAACTCTTTACCCAGGGGATGCAACGTATTACTAGCGGTCTGGGTAATGTGGGT 537
Db    2450 GACGATTTAATCACTAATGCTGTTGCCAGATTAAATAAATTTCTTGGTGCAGCGGGT 2506


RESULT   8
W10PPIBCA
LOCUS       W10PPIBCA    4761 bp    DNA             BCT       07-MAY-1996
DEFINITION  Plasmid pWR100 (from S.flexneri) antigen ippI, ipaB, ipaC genes,
            complete cds, and ippD gene, 3' end.
ACCESSION   M34849
NID         g155327
VERSION     M34849.1  GI:155327
KEYWORDS    antigen.
SOURCE      Plasmid pWR100 (clone: HB101/pHS[4108,5357].).
  ORGANISM  Plasmid pWR100
            plasmids.
REFERENCE   1  (bases 1 to 4761; 1 to 538; 3987 to 4761)
  AUTHORS   Baudry,B., Kaczorek,M. and Sansonetti,P.J.
  TITLE     Nucleotide sequence of the invasion plasmid antigen B and C genes
            (ipaB and ipaC) of Shigella flexneri
  JOURNAL   Microb. Pathog. 4 (5), 345-357 (1988)
  MEDLINE   89200844
COMMENT     Draft entry and computer-readable sequence for [1] kindly submitted
            by B.Baudry, 05-JUN-1990.
               Center for Vaccine Development
               University of Maryland at Baltimore
               10 South Pine Street
               Baltimore, MD 21201.
FEATURES             Location/Qualifiers
     source          1. .4761
                     /organism="Plasmid pWR100"
                     /plasmid="Plasmid pWR100"
                     /specific_host="M90T"
                     /db_xref="taxon:2661"
                     /clone="HB101/pHS[4108,5357]."
     CDS             <1. .549
                     /note="ORF"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA98422.1"
                     /db_xref="PID:g155328"
                     /db_xref="GI:155328"
                     /translation="KLVKKILSVFNLKQRFPQKNFGCPVNINKIRDSVIDKIKDSNSG
                     NQLFCWMSQERTTYVSSMINRSIDEMAIHNGVVLTSDNKKNIFAAIEKKFPDIKLDEK
                     SAQTSISHTALNEIASSGLRAKILKRYSSDMDLFNTQMKDLTNLVSSSVYDKIFNEST
                     KVLQIEISAEVLKAVYRQSNTN"
     -10_signal      556. .561
                     /gene="ippI"
     gene            556. .1073
                     /gene="ippI"
     RBS             595. .600
                     /gene="ippI"
     CDS             606. .1073
                     /gene="ippI"
                     /codon_start=1
                     /transl_table=11
                     /product="antigen"
                     /protein_id="AAA98423.1"
                     /db_xref="PID:g155329"
                     /db_xref="GI:155329"
                     /translation="MSLNITENESISTAVIDAINSGATLKDINAIPDDMMDDIYSYAY
                     DFYNKGRIEEAEVFFRFLCIYDFYNVDYIMGLAAIYQIKEQFQQAADLYAVAFALGKN
                     DYTPVFHTGQCQLRLKAPLKAKECFELVIQHSNDEKLKIKAQSYLDAIQDIKE"
     gene            1065. .2821
                     /gene="ipaB"
     RBS             1065. .1070
                     /gene="ipaB"
     CDS             1079. .2821
                     /gene="ipaB"
                     /codon_start=1
                     /transl_table=11
                     /product="antigen"
                     /protein_id="AAA98424.1"
                     /db_xref="PID:g155330"
                     /db_xref="GI:155330"
                     /translation="MHNVSTTTTGFPLAKILTSTELGDNTIQAANDAANKLFSLTIAD
                     LTANQNINTTNAHSTSNILIPELKAPKSLNASSQLTLLIGNLIQILGEKSLTALTNKI
                     TAWKSQQQARQQKNLEFSDKINTLLSETEGLTRDYEKQINKLKNADSKIKDLENKINQ
                     IQTRLSELDPESPEKKKLSREEIQLTIKKDAAVKDRTLIEQKTLSIHSKLTDKSMQLE
                     KEIDSFSAFSNTASAEQLSTQQKSLTGLASVTQLMATFIQLVGKNNEESLKNDLALFQ
                     SLQESRKTEMERKSDEYAAEVRKAEELNRVMGCVGKILGALLTIVSVVAAAFSGGASL
                     ALAAVGLALMVTDAIVQAATGNSFMEQALNPIMKAVIEPLIKLLSDAFTKMLEGLGVD
                     SKKAKMIGSILGAIAGALVLVAAVVLVATVGKQAAAKLAENIGKIIGKTLTDLIPKFL
                     KNFSSQLDDLITNAVARLNKFLGAAGDEVISKQIISTHLNQAVLLGESVNSAIQAGGS
                     VASAVFQNSASTNLADLTLSKYQVEQLSKYISEAIEKFGQLQEVIADLLASMSNSQAN
                     RTDVAKAILQQTTA"
     gene            2723. .3932
                     /gene="ipaC"
     -35_signal      2723. .2728
                     /gene="ipaC"
     -10_signal      2745. .2750
                     /gene="ipaC"
     RBS             2829. .2834
                     /gene="ipaC"
     CDS             2841. .3932
                     /gene="ipaC"
                     /codon_start=1
                     /transl_table=11
                     /product="antigen"
                     /protein_id="AAA98425.1"
                     /db_xref="PID:g155331"
                     /db_xref="GI:155331"
                     /translation="MEIQNTKPTQTLYTDISTKQTQSSSETQKSQNYQQIAAHIPLNV
                     GKNPVLTTTLNDDQLLKLSEQVQHDSEIIARLTDKKMKDLSEMSHTLTPENTLDISSL
                     SSNAVSLIISVAVLLSALRIAETKLGSQLSLIAFDATKSAAENIVRQGLAALSSSITG
                     AVTQVGITGIGAKKTHSGISDQKGALRKNLATAQSLEKELAGSKLGLNKQIDTNITSP
                     QTNSSTKFLGKNKLAPDNISLSTEHKTSLSSPDISLQDKIDTQRRTYELNTLSAQQKQ
                     NIGRATMETSAVAGNISTSGGRYASALEEEEQLISQASSKQAEEASQVSKEASQATNQ
```

```
                 LIQKLLNIIDSINQSKNSAASQIAGNIRA"
     RBS         3971. .3977
                 /gene="ipaD"
     gene        3971. .4761
                 /gene="ipaD"
     CDS         3983. .>4761
                 /gene="ipaD"
                 /codon_start=1
                 /transl_table=11
                 /product="antigen"
                 /protein_id="AAA98426.1"
                 /db_xref="PID:g552045"
                 /db_xref="GI:552045"
                 /translation="MNITTLTNSISTSSFSPNNTNGSSTETVNSDIKTTTSSHPVSSL
                 TMLNDTLHNIRTTNQALKKELSQKTLTKTSLEEIALHSSQISMDVNKSAQLLDILSRN
                 EYPINKDARELLHSAPKEAELDGDQMISHRELWAKIANSINDINEQYLKVYEHAVSSY
                 TQMYQDFSAVLSSLAGWISPGGNDGNSVKLQVNSLKKALEELKEKYKDKPLYPANNTV
                 SQEQANKWLTELGGTIGKVSQKNGGYVVSINMTPIDNMLKS"
BASE COUNT     1769 a    952 c    805 g   1235 t
ORIGIN
```

```
  Query Match             20.5%;  Score 178.6;  DB 1;  Length 4761;
  Best Local Similarity   58.3%;  Pred. No. 6.6e-34;
  Matches  313;  Conservative    0;  Mismatches  224;  Indels    0;  Gaps    0;

Qy       1 CGCAAAGCCGAGGAAACGAACCGCATTATGGGATGTATCGGGAAAGTCCTCGGCGCGCTG 60
Db    1970 CGTAAAGCAGAAGAACTCAACAGAGTAATGGGTTGTGTTGGGAAAATACTTGGGGCACTT 2029

Qy      61 CTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCT 120
Db    2030 TTAACTATCGTTAGTGTTGTTGCAGCAGCTTTTTCTGGAGGAGCCTCTCTAGCACTGGCA 2089

Qy     121 GCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCG 180
Db    2090 GCTGTTGGTTTAGCTCTTATGGTTACGGATGCTATAGTACAAGCAGCGACCGGCAATTCC 2149

Qy     181 TTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTG 240
Db    2150 TTCATGGAACAAGCCCTGAATCCGATCATGAAAGCAGTCATTGAACCCTTAATCAAACTC 2209

Qy     241 ATTGGCAAGGCGATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAG 300
Db    2210 CTTTCAGATGCATTTACAAAAATGCTCGAAGGCTTGGGCGTCGACTCGAAAAAAGCCAAA 2269

Qy     301 ATGGCCGGCAGCATTGTTGGTGCGATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTG 360
Db    2270 ATGATTGGCTCTATTCTGGGGGCAATCGCAGGCGCTCTTGTCCTAGTTGCAGCAGTCGTT 2329

Qy     361 GTGGTCGCAGTTGTCGGGAAAGGCGCGGCGGCGAAACTGGGTAACGCGCTGAGCAAAATG 420
Db    2330 CTCGTAGCCACTGTTGGTAAACAGGCAGCAGCAAAACTTGCAGAAAATATTGGCAAAATA 2389

Qy     421 ATGGGCGAAACGATTAAGAAGTTGGTGCCTAACGTGCTGAAACAGTTGGCGCAAAACGGC 480
Db    2390 ATAGGTAAAACCCTCACAGACCTTATACCAAAGTTTCTCAAGAATTTTTCTTCTCAACTG 2449

Qy     481 AGCAAACTCTTTACCCAGGGGATGCAACGTATTACTAGCGGTCTGGGTAATGTGGGT 537
Db    2450 GACGATTTAATCACTAATGCTGTTGCCAGATTAAATAAATTTCTTGGTGCAGCGGGT 2506
```

```
RESULT   9
AC007249
LOCUS       AC007249   207466 bp    DNA              HTG       09-APR-1999
DEFINITION  Homo sapiens clone NH0320M02, WORKING DRAFT SEQUENCE, 5 unordered
            pieces.
ACCESSION   AC007249
NID         g4580499
VERSION     AC007249.2  GI:4580499
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
```

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 207466)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 207466)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-APR-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Apr 9, 1999 this sequence version replaced gi:4572717.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    12220: contig of 12220 bp in length
            *   12221    12238: gap of unknown length
            *   12239    30966: contig of 18728 bp in length
            *   30967    30984: gap of unknown length
            *   30985    50458: contig of 19474 bp in length
            *   50459    50476: gap of unknown length
            *   50477   105779: contig of 55303 bp in length
            *  105780   105797: gap of unknown length
            *  105798   207466: contig of 101669 bp in length.
FEATURES            Location/Qualifiers
     source         1. .207466
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="NH0320M02"
BASE COUNT    51723 a  53391 c  53578 g  48702 t     72 others
ORIGIN
```

```
  Query Match              8.3%;  Score 72.6;  DB 35;  Length 207466;
  Best Local Similarity   47.7%;  Pred. No. 1e-07;
  Matches  213;  Conservative    0;  Mismatches  234;  Indels    0;  Gaps    0;

Qy      25 ATTATGGGATGTATCGGGAAAGTCCTCGGCGCGCTGCTAACCATTGTCAGCGTTGTGGCC 84
Db  135622 ATGGTGTGAATGTGCTGGAAAGTGCCTTCTGATCCCAGAGTCCCTCAACTCGTGGTGGTG 135681

Qy      85 GCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTG 144
Db  135682 GTGGTGGCGGTGGCGGTGGCGGTGGTGGTGGCGGTGGCGGTAGTGGTGGCGGTGATGGTG 135741

Qy     145 GCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCGTTTATTCAGCAGGCGCTAAACCCG 204
Db  135742 GCGGTGGCAGTGGCGGTGGTGGCGGCGGTGGTGGTGATGGTGATGGTGGCGGTGATGGTG 135801

Qy     205 ATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTGATTGGCAAGGCGATTACCAAAGCG 264
Db  135802 GTGATGATGGTGATGGTGGTGGTGGTGGTGATGGTGATGGTGGTGGTGGTGGTGATGGTG 135861

Qy     265 CTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAGATGGCCGGCAGCATTGTTGGTGCG 324
Db  135862 GTGGTGGTGGTGGTGGTGGTGATGGTGGTGGTGGTGGTGGTGGTGGTGATGATGGTAGTG 135921

Qy     325 ATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTGGTGGTCGCAGTTGTCGGGAAAGGC 384
Db  135922 GTGGTGATGGTGGTGGTGGTGGTGGTGGTGGTGATGGTGGTGGTGGTGGTGGTGGTGGTG 135981

Qy     385 GCGGCGGCGAAACTGGGTAACGCGCTGAGCAAAATGATGGGCGAAACGATTAAGAAGTTG 444
Db  135982 ATGGTGGTGGTGGTGGTGATGGTGGTGGTGGTGGTGATGGTGGTGGTGGTGGTGGTGATG 136041

Qy     445 GTGCCTAACGTGCTGAAACAGTTGGCG 471
```

```
Db 136042 GTGGTGGTGGTGGTGGTGGTGGTGGTG 136068


RESULT 10
AC007009
LOCUS       AC007009   158880 bp    DNA             HTG       05-MAR-1999
DEFINITION  Homo sapiens clone NH0560C01, WORKING DRAFT SEQUENCE, 3 unordered
            pieces.
ACCESSION   AC007009
NID         g4337223
VERSION     AC007009.1  GI:4337223
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 158880)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 158880)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAR-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     6180: contig of 6180 bp in length
            *     6181     6198: gap of unknown length
            *     6199    40395: contig of 34197 bp in length
            *    40396    40413: gap of unknown length
            *    40414   158880: contig of 118467 bp in length.
FEATURES             Location/Qualifiers
     source          1. .158880
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="NH0560C01"
BASE COUNT    48816 a  32181 c  31490 g  46357 t     36 others
ORIGIN


  Query Match             8.3%;  Score 72;  DB 34;  Length 158880;
  Best Local Similarity   49.5%;  Pred. No. 1.4e-07;
  Matches  186;  Conservative    0;  Mismatches  190;  Indels    0;  Gaps    0;

Qy      25 ATTATGGGATGTATCGGGAAAGTCCTCGGCGCGCTGCTAACCATTGTCAGCGTTGTGGCC 84
Db     667 ATGGTGGTAGTGGTGGGGGTAGTGGTGATGAAGGAGATAATGGTGGTAGTGGTGGTGGAG 726

Qy      85 GCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTG 144
Db     727 GTGATGGTGATGGTGGCGGAGGAGGTGATGGTGGTAGTGGTGGGGGTAGTGGTGTTGAAG 786

Qy     145 GCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCGTTTATTCAGCAGGCGCTAAACCCG 204
Db     787 GAGATGGTGGTGGTGGTGGTGGAGGTGATGGTGATGGTGATAGTGGTGGGGGTAGTTGTG 846

Qy     205 ATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTGATTGGCAAGGCGATTACCAAAGCG 264
Db     847 ATGAAGGAGATGGTGGTTGTGGTGGTGGTGGTGATGGTGGAGGAGGTGATGGTGGTAGTG 906

Qy     265 CTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAGATGGCCGGCAGCATTGTTGGTGCG 324
Db     907 GTGGGGGTAGTGGTGATTAAGGAGATAATGGTGGTGGTGGTTGTGGTTGTTGTGGAGGAG 966

Qy     325 ATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTGGTGGTCGCAGTTGTCGGGAAAGGC 384
Db     967 ATGGTGGTTGTTGTGGAAGAGGTGATGGTGGTAGTGCTGGGGGTAGTGGTGGTGAAGGAG 1026

Qy     385 GCGGCGGCGAAACTGG 400
Db    1027 ATGGTGGTGGTGGTGG 1042


RESULT 11
AC004928/c
LOCUS       AC004928   137137 bp    DNA             PRI       17-APR-1999
DEFINITION  Homo sapiens PAC clone DJ0914M06 from 7q33-q35, complete sequence.
ACCESSION   AC004928
NID         g4587636
VERSION     AC004928.2  GI:4587636
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 137137)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
REFERENCE   2  (bases 1 to 137137)
  AUTHORS   Geisel,C. and Cloud,J.
  TITLE     The sequence of Homo sapiens PAC clone DJ0914M06
  JOURNAL   Unpublished (1999)
REFERENCE   3  (bases 1 to 137137)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-1998) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 137137)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-APR-1999) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Apr 19, 1999 this sequence version replaced gi:3213069.
            SUBMITTED BY:  WUGSC
            Genome Sequencing Center
            Department of Genetics
            Washington University
            St. Louis MO 63108, USA
            http://genome.wustl.edu/gsc
            mailto:sapiens@watson.wustl.edu

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            The sequence of this clone was established as part of a mapping and
            sequencing collaboration between the NHGRI Chromosome 7 Mapping
            Project (Eric D. Green, Director), John D. McPherson in the
            Department of Genetics (Washington University), and the Washington
            University Genome Sequencing Center.  For additional information
            about the map position of this sequence, see
            http://www.nhgri.nih.gov/DIR/GTB/CHR7 , send
            mailto:egreen@nhgri.nih.gov , or see http://genome.wustl.edu/gsc
```

SOURCE INFORMATION:
This clone was derived from human PAC library RPCI-5, prepared by Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://bacpac.med.buffalo.edu) using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor.
The clone may be obtained either from Genome Systems, Inc. (http://www.genomesystems.com) or Research Genetics, Inc. (http://www.resgen.com); or from Pieter de Jong.
VECTOR: pCYPAC2
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is DJ0733H01, 200 bp overlap; the clone sequenced to the right is DJ0842K16, 200 bp overlap. Actual start of this clone is at base position 197 of DJ0914M06; actual end is at 136941 of DJ0914M06.

There is a tandem repeat within this sequence at base positions 79937 to 80547. The tandem consists of TC rich sequence and there was a forced join between two clones at base position 80183. Based on the HindIII digest and the EcoRV digest the assembly looks correct.

FEATURES	Location/Qualifiers
source	1. .137137
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="7"
	/map="7q33-q35"
	/clone_lib="RPCI-5"
	/clone="DJ0914M06"
repeat_region	80. .439
	/rpt_family="L2"
repeat_region	917. .958
	/rpt_family="AT_rich"
repeat_region	959. .1459
	/rpt_family="L1"
repeat_region	1460. .1651
	/rpt_family="MER1_type"
repeat_region	1652. .3086
	/rpt_family="L1"
repeat_region	3913. .3947
	/rpt_family="AT_rich"
repeat_region	5654. .5696
	/rpt_family="(TA)n"
repeat_region	6993. .7548
	/rpt_family="MER21_g"
repeat_region	8083. .8315
	/rpt_family="MIR"
repeat_region	8738. .8817
	/rpt_family="L2"
repeat_region	11249. .12279
	/rpt_family="Mariner"
repeat_region	12280. .12586
	/rpt_family="Alu"
repeat_region	12587. .12824
	/rpt_family="Mariner"
repeat_region	12975. .13291
	/rpt_family="Alu"
repeat_region	13536. .13600
	/rpt_family="Alu"
misc_feature	14686. .15096
	/note="match to EST M78298 (NID:g272614)"
STS	complement(14686. .15096)
	/db_xref="GI:3660601"
repeat_region	14909. .15105
	/rpt_family="MIR"
repeat_region	15980. .16638
	/rpt_family="L1"
repeat_region	16981. .17519
	/rpt_family="L2"
repeat_region	18243. .18385
	/rpt_family="L1"
misc_feature	18883. .19159
	/note="match to EST T05081 (NID:g316233)"
repeat_region	19476. .19522
	/rpt_family="MER2_type"
repeat_region	19523. .19901
	/rpt_family="MaLR"
repeat_region	19902. .20807
	/rpt_family="MER2_type"
repeat_region	20806. .20907
	/rpt_family="MER2_type"
repeat_region	21492. .21803
	/rpt_family="Alu"
repeat_region	23485. .23843
	/rpt_family="MaLR"
repeat_region	23941. .24048
	/rpt_family="MER1_type"
repeat_region	24852. .25345
	/rpt_family="L2"
STS	complement(25557. .25753)
	/db_xref="GI:3800469"
misc_feature	25559. .25753
	/note="match to EST AA076702 (NID:g1836190)"
repeat_region	26221. .26272
	/rpt_family="MIR"
misc_feature	26393. .26597
	/note="match to EST AA077210 (NID:g1836684)"
repeat_region	28261. .28342
	/rpt_family="L1"
repeat_region	28343. .28649
	/rpt_family="Alu"
repeat_region	28650. .28675
	/rpt_family="L1"
repeat_region	28756. .29288
	/rpt_family="L1"
misc_feature	29340. .29573
	/note="match to EST AA338345 (NID:g1990593)"
repeat_region	30247. .30336
	/rpt_family="purine-rich"
repeat_region	31667. .31726
	/rpt_family="A-rich"
repeat_region	32386. .32487
	/rpt_family="MIR"
repeat_region	32626. .32936
	/rpt_family="Alu"
repeat_region	33045. .33193
	/rpt_family="Alu"
repeat_region	33712. .33804
	/rpt_family="(CA)n"
repeat_region	35827. .36129
	/rpt_family="Alu"
misc_feature	36930. .37225
	/note="match to EST AA076689 (NID:g1836317)"
STS	36930. .37225
	/db_xref="GI:3800536"
repeat_region	37217. .37318
	/rpt_family="MIR"
repeat_region	39267. .39559
	/rpt_family="Alu"
repeat_region	39627. .39730
	/rpt_family="L2"
repeat_region	39814. .39942
	/rpt_family="L1"
repeat_region	40542. .40566
	/rpt_family="(TAA)n"
repeat_region	40567. .41589
	/rpt_family="L1"
repeat_region	42368. .42564

Query Match 8.2%; Score 71.4; DB 42; Length 137137;
Best Local Similarity 47.6%; Pred. No. 1.9e-07;
Matches 210; Conservative 0; Mismatches 231; Indels 0; Gaps 0;

```
Qy     37 ATCGGGAAAGTCCTCGGCGCGCTGCTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACC 96
          |||  |   ||  | || | | || |     | ||     | ||||  |  ||  |
Db  80443 ATCATGGTGGTAGTGGGGGTGGTGGTGGTGGTGGTGATGGTGGTGGTGGTGATGGTG 80384
```

```
Qy      97 GGTGGGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATT 156
Db   80383 GTGGTGCTGGTGGTGGTGGTGGTGGTGGTGATGGTGGTGGTGATGATGGTGGTGGTG 80324

Qy     157 GTGAAGGCGGCGACGGGAGTGTCGTTTATTCAGCAGGCGCTAAACCCGATTATGGAGCAT 216
Db   80323 GTGATGGTGGTGATGGTGGTGGTGGTGATGATGATGGTAGTGATGGGGGTGGTGGTGGTG 80264

Qy     217 GTGCTGAAGCCGTTAATGGAGCTGATTGGCAAGGCGATTACCAAAGCGCTGGAAGGATTA 276
Db   80263 GTAGTGATGGTGGTGGTGGTGATGGTGGTGGTGGTGGTGGTGGTGATGGTGGTGATGGTG 80204

Qy     277 GGCGTCGATAAGAAAACGGCAGAGATGGCCGGCAGCATTGTTGGTGCGATTGTCGCCGCT 336
Db   80203 GTGGTGGTGATGATGATGGTAGTGATGGGGGTGGTGGTGGTGGTGGTAGTGATGGTGGTG 80144

Qy     337 ATTGCCATGGTGGCGGTCATTGTGGTGGTCGCAGTTGTCGGGAAAGGCGCGGCGGCGAAA 396
Db   80143 GTGGTGATGGTGGTGGTGATGGTGGTGGTGGTGGTGGTGATGATGGTGGTGGTGGTA 80084

Qy     397 CTGGGTAACGCGCTGAGCAAAATGATGGGCGAAACGATTAAGAAGTTGGTGCCTAACGTG 456
Db   80083 GCGGTGATGGTGGTGGTGGTGATGATGATGGTAGTGATGGTAATGGTTATGATGGTGGTG 80024

Qy     457 CTGAAACAGTTGGCGCAAAAC 477
Db   80023 ATGGTGGTGGTGGTGCTGATC 80003


RESULT  12
AC007263
LOCUS       AC007263   168844 bp    DNA             HTG       24-MAY-1999
DEFINITION  Homo sapiens chromosome 14 clone BAC 79J20 map 14q31, WORKING DRAFT
            SEQUENCE, 5 ordered pieces.
ACCESSION   AC007263
NID         g4885690
VERSION     AC007263.3  GI:4885690
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 168844)
  AUTHORS   Rowen,L., Madan,A., Qin,S., Abbasi,N., Dors,M., Dickhoff,R.,
            James,R., Loretz,C., Lasky,S., Madan,A., Prescott,S., Ratcliffe,A.,
            Shaffer,T. and Hood,L.
  TITLE     Sequencing of human chromosome 14q31 region
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 168844)
  AUTHORS   Rowen,L., Madan,A., Qin,S., Abbasi,N., Dors,M., Dickhoff,R.,
            James,R., Loretz,C., Lasky,S., Madan,A., Prescott,S., Ratcliffe,A.,
            Shaffer,T. and Hood,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-APR-1999) Multimegabase Sequencing Center, University
            of Washington, PO BOX 357730, Seattle, WA 98195, USA
COMMENT     On May 24, 1999 this sequence version replaced gi:4680438.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved
            *        1    23725: contig of 23725 bp in length
            *                    gap of unknown length
            *    23726    34192: contig of 10467 bp in length
            *                    gap of unknown length
            *    34193    50560: contig of 16368 bp in length
            *                    gap of unknown length
            *    50561    78056: contig of 27496 bp in length
            *                    gap of unknown length
            *    78057   168844: contig of 90788 bp in length.
FEATURES             Location/Qualifiers
     source          1. .168844
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="14"
                     /clone="BAC 79J20"
                     /map="14q31"
                     /clone_lib="RPCI-11"
                     /note="This clone contains part of the gene for human
                     checkpoint suppressor 1 (U68723)"
BASE COUNT    42649 a  36845 c  39967 g  49381 t      2 others
ORIGIN


  Query Match             8.1%;  Score 70.2;  DB 35;  Length 168844;
  Best Local Similarity   48.4%;  Pred. No. 3.9e-07;
  Matches  195;  Conservative    0;  Mismatches  208;  Indels    0;  Gaps    0;

Qy      76 GTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTGCG 135
Db   89547 GTGGTGGTGGTGGAGGTGGTGGTGGAGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGTG 89606

Qy     136 GTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCGTTTATTCAGCAGGCG 195
Db   89607 GTGGTGGTGGTGGAGATGGTGGTGGAGGTGGTGGTGGAGATGGTGGTGGTGGTGGTGGTG 89666

Qy     196 CTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTGATTGGCAAGGCGATT 255
Db   89667 GTGGTGGTGGTGGTGGAGATGGTGGTGGAGGTGGTGGTGGAGATGGTGGTGGTGGTGGTG 89726

Qy     256 ACCAAAGCGCTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAGATGGCCGGCAGCATT 315
Db   89727 GAGATGGTGGTGGAGGTGGTGGTGGTCGTAGAGATGGTGGTGGTGGTGGTGGTCGTAGAG 89786

Qy     316 GTTGGTGCGATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTGGTGGTCGCAGTTGTC 375
Db   89787 ATGGTGGTGGTGGTGGTGGTGGAGGTGGTGGTGGAGATGGTGGTGGTGGTCGTAGAG 89846

Qy     376 GGGAAAGGCGCGGCGGCGAAACTGGGTAACGCGCTGAGCAAAATGATGGGCGAAACGATT 435
Db   89847 ATGGTGGTGGTGGTGGTGGAGATGGTGGTGGTGGTGGTGGTGGTGGTGGAGGAGGAGGTG 89906

Qy     436 AAGAAGTTGGTGCCTAACGTGCTGAAACAGTTGGCGCAAAACG 478
Db   89907 GTGGAGCTGGTGCTGGAGGTGGTGGTGGTGGTGGTGGAGGAGG 89949


RESULT  13
AC000097/c
LOCUS       AC000097   162269 bp    DNA             PRI       01-DEC-1998
DEFINITION  Homo sapiens Chromosome 22q11.2 PAC Clone p201m18 In DGCR Region,
            complete sequence.
ACCESSION   AC000097
NID         g3935200
VERSION     AC000097.1  GI:3935200
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 162269)
  AUTHORS   Hua,A. and Roe,B.A.
  TITLE     Homo sapiens Chromosome 22q11.2 PAC Clone p201m18 In DGCR Region
  JOURNAL   Unpublished (1996)
REFERENCE   2  (bases 1 to 162269)
  AUTHORS   Budarf,M.L. and Emanuel,B.S.
  JOURNAL   Unpublished (1996)
REFERENCE   3  (bases 1 to 162269)
  AUTHORS   Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-1996) Department Of Chemistry And Biochemistry,
```

```
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   4  (bases 1 to 162269)
  AUTHORS   Hua,A., Budarf,M.L., Emanuel,B.S. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-FEB-1998) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   5  (bases 1 to 162269)
  AUTHORS   Hua,A., Budarf,M.L., Emanuel,B.S. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-FEB-1998) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   6  (bases 1 to 162269)
  AUTHORS   Hua,A., Budarf,M.L., Emanuel,B.S. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-NOV-1998) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   7  (bases 1 to 162269)
  AUTHORS   Hua,A., Budarf,M.L., Emanuel,B.S. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-DEC-1998) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On Dec 1, 1998 this sequence version replaced gi:3845359.
FEATURES             Location/Qualifiers
     source          1. .162269
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="22"
                     /map="22q11.2"
                     /clone="p201m18"
BASE COUNT    31852 a  46339 c  47075 g  37003 t
ORIGIN


  Query Match             7.8%;  Score 67.8;  DB 11;  Length 162269;
  Best Local Similarity   47.0%;  Pred. No. 1.5e-06;
  Matches  210;  Conservative    0;  Mismatches  237;  Indels    0;  Gaps    0;

Qy      25 ATTATGGGATGTATCGGGAAAGTCCTCGGCGCGCTGCTAACCATTGTCAGCGTTGTGGCC 84
           ||  |||       | | |  |||  | |  | | || |      |  |    || ||||
Db 111986 ATGGTGGTGGTGGTGGTGGTAGTGATGGTGGTGGTGGTGGTAGTGATGGTGGTGGTGGTG 111927

Qy      85 GCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTG 144
           |  ||  | |  |  | || |    ||| | |||  | | |||   | | |||  || ||
Db 111926 GTGGTAGTGATGGTGGTGGTGGTGGTGGTGGTGGTAGTGATGGTGGTGGTGGTGGTGATG 111867

Qy     145 GCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCGTTTATTCAGCAGGCGCTAAACCCG 204
           |  |  |   | |||| || |  |  ||  |||  | |  |  |  || | | |    |
Db 111866 GTGGTGGTGGTAGTGATGGTGATGGTGGTGGTGGTGATGGTGGTGGTGGTGGTGATGGTG 111807

Qy     205 ATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTGATTGGCAAGGCGATTACCAAAGCG 264
           |  ||  |    || ||  |  | |  | | | || | |    || | |       || |
Db 111806 GTAGTGATGGTGATGGTGGTGGTGGTGGTAGTGATGGTGGTGGTGGTGGTGGTGGTAGTG 111747

Qy     265 CTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAGATGGCCGGCAGCATTGTTGGTGCG 324
           |||  |   | |  || |  | |      ||  | | |||  |  |   | || |  | |
Db 111746 ATGGTGGTGGTGGTGGTAGTGATGGTGGTGGTGGTGGTAGTGATGGTGGTGGTGGTG 111687

Qy     325 ATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTGGTGGTCGCAGTTGTCGGGAAAGGC 384
           | || |  |  || |  ||||| ||| || |||||||| | |||  | | |   |
Db 111686 GTGGTGGTAGTGATGGTGGTGGTGGTGGTGATGGTGGTGGTGGTAGTGATGGTGGTGGTG 111627

Qy     385 GCGGCGGCGAAACTGGGTAACGCGCTGAGCAAAATGATGGGCGAAACGATTAAGAAGTTG 444
           | || ||    |||     | | ||     | ||||||  |    | |  || | ||
Db 111626 GTGGTGGTAGTGATGGTGGTGGTGGTGGTGGTAGTGATGGTGGTGGTGGTGGTGATGGTG 111567

Qy     445 GTGCCTAACGTGCTGAAACAGTTGGCG 471
           |||       ||  |||    | ||| |
Db 111566 GTGGTGGTAGTAGTGATGGTGGTGGTG 111540
```

```
RESULT  14
AC006549/c
LOCUS       AC006549   165321 bp    DNA             HTG       18-MAY-1999
DEFINITION  Homo sapiens, WORKING DRAFT SEQUENCE, 8 unordered pieces.
ACCESSION   AC006549
NID         g4835949
VERSION     AC006549.17  GI:4835949
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 165321)
  AUTHORS   Hu,P., Chen,F., Dumanski,J. and Roe,B.A.
  TITLE     Homo sapiens Chromosome 22q11 PAC Clone p215k21 Distal To DGCR
            Region
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 165321)
  AUTHORS   Hu,P., Chen,F., Dumanski,J. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-FEB-1999) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On May 16, 1999 this sequence version replaced gi:4803977.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     4176: contig of 4176 bp in length
            *     4177     4245: gap of unknown length
            *     4246     6464: contig of 2219 bp in length
            *     6465     6533: gap of unknown length
            *     6534    13373: contig of 6840 bp in length
            *    13374    13442: gap of unknown length
            *    13443    24265: contig of 10823 bp in length
            *    24266    24334: gap of unknown length
            *    24335    36267: contig of 11933 bp in length
            *    36268    36335: gap of unknown length
            *    36336    52372: contig of 16037 bp in length
            *    52373    52440: gap of unknown length
            *    52441    68271: contig of 15831 bp in length
            *    68272    68339: gap of unknown length
            *    68340   165321: contig of 96982 bp in length.
FEATURES             Location/Qualifiers
     source          1. .165321
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT    32764 a  49257 c  48950 g  33869 t    481 others
ORIGIN


  Query Match             7.8%;  Score 67.8;  DB 35;  Length 165321;
  Best Local Similarity   47.0%;  Pred. No. 1.5e-06;
  Matches  210;  Conservative    0;  Mismatches  237;  Indels    0;  Gaps    0;

Qy      25 ATTATGGGATGTATCGGGAAAGTCCTCGGCGCGCTGCTAACCATTGTCAGCGTTGTGGCC 84
           ||  |||       | | |  |||  | |  | | || |      |  |    || ||||
Db 110322 ATGGTGGTGGTGGTGGTGGTAGTGATGGTGGTGGTGGTGGTAGTGATGGTGGTGGTGGTG 110263

Qy      85 GCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTG 144
           |  ||  | |  |  | || |    ||| | |||  | | |||   | | |||  || ||
Db 110262 GTGGTAGTGATGGTGGTGGTGGTGGTGGTGGTGGTAGTGATGGTGGTGGTGGTGGTGATG 110203

Qy     145 GCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCGTTTATTCAGCAGGCGCTAAACCCG 204
           |  |  |   | |||| || |  |  ||  |||  | |  |  |  || | | |    |
Db 110202 GTGGTGGTGGTAGTGATGGTGATGGTGGTGGTGGTGATGGTGGTGGTGGTGGTGATGGTG 110143
```

```
Qy    205 ATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTGATTGGCAAGGCGATTACCAAAGCG 264
Db 110142 GTAGTGATGGTGATGGTGGTGGTGGTGGTAGTGATGGTGGTGGTGGTGGTGGTGGTAGTG 110083

Qy    265 CTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAGATGGCCGGCAGCATTGTTGGTGCG 324
Db 110082 ATGGTGGTGGTGGTGGTAGTGATGGTGGTGGTGGTGGTGGTAGTGATGGTGGTGGTGGTG 110023

Qy    325 ATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTGGTGGTCGCAGTTGTCGGGAAAGGC 384
Db 110022 GTGGTGGTAGTGATGGTGGTGGTGGTGGTGATGGTGGTGGTGGTAGTGATGGTGGTGGTG 109963

Qy    385 GCGGCGGCGAAACTGGGTAACGCGCTGAGCAAAATGATGGGCGAAACGATTAAGAAGTTG 444
Db 109962 GTGGTGGTAGTGATGGTGGTGGTGGTGGTGGTAGTGATGGTGGTGGTGGTGGTGATGGTG 109903

Qy    445 GTGCCTAACGTGCTGAAACAGTTGGCG 471
Db 109902 GTGGTGGTAGTGATGGTGGTGGTGGTG 109876


RESULT 15
AC006547/c
LOCUS       AC006547   179269 bp    DNA             PRI       20-APR-1999
DEFINITION  , complete sequence.
ACCESSION   AC006547
NID         g4589984
VERSION     AC006547.9  GI:4589984
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 179269)
  AUTHORS   Hu,P., Chen,F., Dumanski,J. and Roe,B.A.
  TITLE     Homo sapiens Chromosome 22q11 PAC Clone p158119 Distal To DGCR
            Region
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 179269)
  AUTHORS   Hu,P., Chen,F., Dumanski,J. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-FEB-1999) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   3  (bases 1 to 179269)
  AUTHORS   Hu,P., Chen,F., Dumanski,J. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-MAR-1999) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   4  (bases 1 to 179269)
  AUTHORS   Hu,P., Chen,F., Dumanski,J. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-1999) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On Apr 20, 1999 this sequence version replaced gi:4454503.
FEATURES             Location/Qualifiers
     source          1. .179269
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT    35548 a  50836 c  51499 g  41386 t
ORIGIN


  Query Match             7.8%;  Score 67.8;  DB 42;  Length 179269;
  Best Local Similarity   47.0%;  Pred. No. 1.5e-06;
  Matches  210;  Conservative    0;  Mismatches  237;  Indels    0;  Gaps    0;

Qy     25 ATTATGGGATGTATCGGGAAAGTCCTCGGCGCGCTGCTAACCATTGTCAGCGTTGTGGCC 84
Db 124204 ATGGTGGTGGTGGTGGTGGTAGTGATGGTGGTGGTGGTGGTAGTGATGGTGGTGGTGGTG 124145

Qy     85 GCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTG 144
Db 124144 GTGGTAGTGATGGTGGTGGTGGTGGTGGTGGTAGTGATGGTGGTGGTGGTGGTGATG 124085

Qy    145 GCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCGTTTATTCAGCAGGCGCTAAACCCG 204
Db 124084 GTGGTGGTGGTAGTGATGGTGATGGTGGTGGTGGTGATGGTGGTGGTGGTGGTGATGGTG 124025

Qy    205 ATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTGATTGGCAAGGCGATTACCAAAGCG 264
Db 124024 GTAGTGATGGTGATGGTGGTGGTGGTGGTAGTGATGGTGGTGGTGGTGGTGGTGGTAGTG 123965

Qy    265 CTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAGATGGCCGGCAGCATTGTTGGTGCG 324
Db 123964 ATGGTGGTGGTGGTGGTAGTGATGGTGGTGGTGGTGGTGGTAGTGATGGTGGTGGTGGTG 123905

Qy    325 ATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTGGTGGTCGCAGTTGTCGGGAAAGGC 384
Db 123904 GTGGTGGTAGTGATGGTGGTGGTGGTGGTGATGGTGGTGGTGGTAGTGATGGTGGTGGTG 123845

Qy    385 GCGGCGGCGAAACTGGGTAACGCGCTGAGCAAAATGATGGGCGAAACGATTAAGAAGTTG 444
Db 123844 GTGGTGGTAGTGATGGTGGTGGTGGTGGTGGTAGTGATGGTGGTGGTGGTGGTGATGGTG 123785

Qy    445 GTGCCTAACGTGCTGAAACAGTTGGCG 471
Db 123784 GTGGTGGTAGTGATGGTGGTGGTGGTG 123758


Search completed: October  5, 1999, 17:55:59
Job time: 11202 sec
```

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 1999, 17:14:06 ; Search time 1495.71 Seconds
(without alignments)
2028.305 Million cell updates/sec

Title: US-09-068-804-13
Perfect score: 1538
Sequence: 1 ATGCGTGATTGCCTGAATAA.....CCCGGTGGCGTCTTAACTGA 1538

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 986266752 residues

Database : EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	39.8	2.6	342	22	R71694	R71694 yj85b03.r1
	2	39	2.5	280	20	M79057	M79057 EST01205 Su
c	3	38	2.5	367	37	AA721338	AA721338 nz73b02.s
c	4	37.2	2.4	553	45	AI338945	AI338945 qg29a11.x
	5	37	2.4	549	41	AI006602	AI006602 ue15g11.y
c	6	36.6	2.4	474	44	AI305207	AI305207 ql90d04.x
	7	36.2	2.4	373	36	C69795	C69795 C69795 Yuji
	8	35.8	2.3	360	36	C67933	C67933 C67933 Yuji
	9	35.8	2.3	376	36	C69422	C69422 C69422 Yuji
	10	35.6	2.3	360	24	D69334	D69334 CELK068F1F
c	11	35.6	2.3	647	46	AI402251	AI402251 GH11218.3
	12	35.4	2.3	547	49	AI660752	AI660752 we66h08.x
	13	35.2	2.3	353	24	D61630	D61630 HUM424H04B
	14	35.2	2.3	403	44	AU045269	AU045269 AU045269
c	15	35	2.3	494	44	AI287656	AI287656 ql53a03.x
	16	34.6	2.2	327	21	T89510	T89510 yd98f07.r1
	17	34.6	2.2	360	24	D69836	D69836 CELK092GYF
c	18	34.6	2.2	377	28	AA072586	AA072586 mm73b05.r
c	19	34.6	2.2	353	48	AI561848	AI561848 vk35f07.y
c	20	34.6	2.2	698	53	HSM007259	A1042409 Homo sapi
	21	34.4	2.2	401	22	R51569	R51569 yg73b04.s1
c	22	34.4	2.2	328	31	AA312178	AA312178 EST182867
	23	34.4	2.2	543	49	AI649600	AI649600 486068B09
c	24	34.2	2.2	500	20	D15747	D15747 RICC1183A R
	25	34.2	2.2	369	25	N42825	N42825 yy12e02.r1
c	26	34.2	2.2	162	26	W23750	W23750 zb46f11.r1
	27	34	2.2	501	22	H10472	H10472 yl90a12.r1
c	28	34	2.2	517	31	AA315321	AA315321 EST187107
c	29	34	2.2	307	32	AA359040	AA359040 EST68073
	30	34	2.2	350	45	AI383400	AI383400 tc97c04.x
	31	33.8	2.2	528	49	AI647511	AI647511 uk40e12.x
	32	33.6	2.2	362	21	T75222	T75222 yc86g07.r1
c	33	33.4	2.2	162	23	H55474	H55474 CHR220413 C
c	34	33.4	2.2	448	26	W93209	W93209 zh45g02.r1
	35	33.4	2.2	542	36	AA614081	AA614081 no94d08.s
	36	33.4	2.2	277	38	AA807096	AA807096 oc31c10.s
c	37	33.4	2.2	324	48	AI562188	AI562188 vw77c08.x
c	38	33.4	2.2	632	51	AI728716	AI728716 BNLGHi114
c	39	33.2	2.2	392	20	T19837	T19837 B702F Heart
	40	33.2	2.2	681	44	AI304873	AI304873 qo19h01.x
	41	33.2	2.2	521	47	AI514261	AI514261 GH27351.5
c	42	33.2	2.2	572	54	HSM010421	A1045571 Homo sapi
c	43	33	2.1	234	20	T28021	T28021 EST24013 Hu
c	44	33	2.1	234	22	R75583	R75583 MDB0014R Mo
	45	33	2.1	434	47	AI478398	AI478398 ti71f07.x

ALIGNMENTS

RESULT 1
R71694/c
LOCUS R71694 342 bp mRNA EST 02-JUN-1995
DEFINITION yj85b03.r1 Soares breast 2NbHBst Homo sapiens cDNA clone IMAGE:155501 5' similar to gb:M88338 SERUM PROTEIN MSE55 (HUMAN);contains MSR1 repetitive element ;, mRNA sequence.
ACCESSION R71694
NID g845726

```
VERSION     R71694.1  GI:845726
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 342)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     On May 9, 1995 this sequence version replaced gi:802985.

            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 1151
            High quality sequence stops: 186
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1151   Std Error: 0.00
            Seq primer: M13RP1
            High quality sequence stop: 186.
FEATURES             Location/Qualifiers
     source          1. .342
                     /organism="Homo sapiens"
                     /db_xref="GDB:573404"
                     /db_xref="taxon:9606"
                     /map="X; 19q13.2"
                     /clone="IMAGE:155501"
                     /clone_lib="Soares breast 2NbHBst"
                     /sex="Female"
                     /dev_stage="adult"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: breast; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of a modified pT7T3 vector (Pharmacia).
                     Library went through one round of normalization to a Cot =
                     230. Library constructed by Bento Soares and M.Fatima
                     Bonaldo."
BASE COUNT       65 a    128 c     97 g     42 t     10 others
ORIGIN
```

```
  Query Match             2.6%;  Score 39.8;  DB 22;  Length 342;
  Best Local Similarity   49.2%;  Pred. No. 0.4;
  Matches   95;  Conservative    0;  Mismatches   98;  Indels    0;  Gaps    0;

Qy      90 TACAACACTGGTTATTCCTGATAATAATCTGACCAGCCTGCCGGAGTTGCCGGAAGGACT 149
           | || | |||| || |  ||      |||  | |   || ||        | |||| |
Db     317 TCCATCTCTGGGTAGTGGTGGCCGCCATCCNAGCCTGCTCCCCAGTNGCTCCGAAGGNCA 258

Qy     150 ACGGGAGCTGGAGGTCTCTGGTAACCTACAACTGACCAGCCTGCCATCGCTGCCGCAGGG 209
             |  ||   |   ||   |    ||| | ||||  | |     || | | ||| | |||
Db     257 GGGCCAGNCAGTCCTCGGGGACCCCCTCCNACTGGGCTGGACTTCACCTCAGCCACTGGG 198

Qy     210 ACTACAGAAGCTGTGGGCCTATAATAATTGGCTGGCCAGCCTGCCGACGTTGCCGCCAGG 269
           | |    ||||||   || ||    |  | |     |  ||  ||  |||| ||  ||
Db     197 CCCAACCCAGCTGTTACCCCATTGGNACAGTGTCCATGGGGTGTGGAGCTTGCGGCAGGG 138

Qy     270 ACTAGGGGATCTG 282
           || ||||||  ||
Db     137 GCTGGGGGATTTG 125
```

```
RESULT   2
M79057
LOCUS       M79057         280 bp    mRNA            EST       26-MAY-1992
DEFINITION  EST01205 Subtracted Hippocampus, Stratagene (cat. #936205) Homo
            sapiens cDNA clone HHCPJ22, mRNA sequence.
ACCESSION   M79057
NID         g273370
VERSION     M79057.1  GI:273370
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 280)
  AUTHORS   Adams,M.D., Dubnick,M., Kerlavage,A.R., Moreno,R., Kelley,J.M.,
            Utterback,T.R., Nagle,J.W., Fields,C. and Venter,J.C.
  TITLE     Sequence identification of 2,375 human brain genes
  JOURNAL   Nature 355, 632-634 (1992)
  MEDLINE   92168112
COMMENT
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            Seq primer: M13 Forward.
FEATURES             Location/Qualifiers
     source          1. .280
                     /organism="Homo sapiens"
                     /db_xref="ATCC (inhost):78579"
                     /db_xref="GDB:D0S2026E"
                     /db_xref="taxon:9606"
                     /clone="HHCPJ22"
                     /clone_lib="Subtracted Hippocampus, Stratagene (cat.
                     #936205)"
                     /note="Vector: lambdaZAP-II; The hippocampus library (#4
                     above) was subtracted with a fibroblast cell line cDNA
                     library (Stratagene cat. #936209; WI38 lung fibroblast
                     cell line; oligo-dT + random primed cDNA synthesis;
                     lambdaZAP-II vector, 1.0kb average insert size.) by the
                     method of Sive & St. John (Nucl. Acids Res. 16:10937,
                     1988)."
BASE COUNT       63 a     67 c     97 g     53 t
ORIGIN
```

```
  Query Match             2.5%;  Score 39;  DB 20;  Length 280;
  Best Local Similarity   58.0%;  Pred. No. 0.6;
  Matches   69;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy    1047 GAACAGATTGCCCGGGAAAAAGCCGGAACACTGGCTTTAGTCGATGACGTTGAGGTCTAT 1106
           || ||| |   || |||| | |  |   ||| || | | || ||||  |||| ||||  |
Db      28 GAGCAGCTCAGCCAGGAAGAGGTTGATGCACAGGTTCTTGTGGATGGTGTTGCGGTCGGT 87

Qy    1107 CTGGCGTTCCAGAATAAGCTGAAGGAATCACTTGAGCTGACCAGCGTGACGTCAGAAAT 1165
           |||  |  || | | |||| |||||      |  || || |||    |||   || ||
Db      88 CTGCAGCCCCCGCAGAAGCAGAAGGTGGAGATGCAGATGGCCAAGCAGACCAGGGAGAT 146
```

```
RESULT   3
AA721338/c
LOCUS       AA721338       367 bp    mRNA            EST       23-JAN-1998
DEFINITION  nz73b02.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1301067 3'
            similar to TR:Q28298 Q28298 RIBOSOME RECEPTOR. ;, mRNA sequence.
ACCESSION   AA721338
NID         g2737473
VERSION     AA721338.1  GI:2737473
KEYWORDS    EST.
```

```
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 367)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1392719.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
            Ph.D., Gerald Marti, M.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Trace considered overall poor quality
            Insert Length: 629   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .367
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1301067"
                     /clone_lib="NCI_CGAP_GCB1"
                     /tissue_type="germinal center B cell"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from human tonsillar cells enriched for
                     germinal center B cells by flow sorting (CD20+, IgD-),
                     provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                     (NCI) and Dr. Gerald Marti (CBER).  cDNA synthesis was
                     primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTTT-
                     3']. Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT       50 a    111 c     96 g    110 t
ORIGIN
```

```
  Query Match             2.5%;  Score 38;  DB 37;  Length 367;
  Best Local Similarity   57.6%;  Pred. No. 1.3;
  Matches   68;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy     953 ACAACTGGTACATAATGCAGAAAAAGGGGAATACGACAACAATCTCCAGGGGCTGGTTTC 1012
           ||| |||    | |    ||||||||| ||| | |  |||||    |  ||||| |   |
Db     167 ACAGCTGAAGGAGAGACTAGAAAAAGGAGAAGAAGTTAACAAGTGACCTGGGCTCGCCGC 108

Qy    1013 CACGGGGCGTGAGATGTTCCGCCTGGCAACACTGGAACAGATTGCCCGGGAAAAAGCC 1070
           |||| | |   ||  | | |    | | || | ||| ||| | ||  |||| || | |
Db     107 CACGAGACTGCAGGAGCTTCTGAAGACGACCCAGGAGCAGCTGGCAAGGGAGAAGGAC 50
```

```
RESULT   4
AI338945/c
LOCUS       AI338945      553 bp    mRNA            EST       13-FEB-1999
DEFINITION  qq29a11.x1 Soares_NhHMPu_S1 Homo sapiens cDNA clone IMAGE:1933916
            3' similar to gb:S78694 PROTEIN-LYSINE 6-OXIDASE PRECURSOR
            (HUMAN);, mRNA sequence.
ACCESSION   AI338945
NID         g4075872
VERSION     AI338945.1  GI:4075872
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 553)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Aug 21, 1998 this sequence version replaced.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 782   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 464.
FEATURES             Location/Qualifiers
     source          1. .553
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1933916"
                     /clone_lib="Soares_NhHMPu_S1"
                     /tissue_type="Pooled human melanocyte, fetal heart, and
                     pregnant uterus"
                     /lab_host="DH10B"
                     /note="Organ: mixed (see below); Vector: pT7T3D-Pac
                     (Pharmacia) with a modified polylinker; Site_1: Not I;
                     Site_2: Eco RI; Equal amounts of plasmid DNA from three
                     normalized libraries (melanocyte 2NbHM, pregnant uterus
                     NbHPU, and fetal heart NbHH19W) were mixed, and ss circles
                     were made in vitro. Following HAP purification, this DNA
                     was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from pools of
                     5,000 clones made from the same 3 libraries. The pools
                     consisted of I.M.A.G.E. clones 260232-265223,
                     340488-345479, and 484488-489479."
BASE COUNT       88 a    166 c    210 g     87 t      2 others
ORIGIN
```

```
  Query Match             2.4%;  Score 37.2;  DB 45;  Length 553;
  Best Local Similarity   53.1%;  Pred. No. 2.8;
  Matches   78;  Conservative    0;  Mismatches   69;  Indels    0;  Gaps    0;

Qy     620 TGGCGGGGCCTTCCGTCCCCGGGAAGCCCGGGCACTGCACCTGGCGGTCGCTGACTGGCT 679
           ||   |||||||     | | |   ||  | ||||||  ||   | | ||| | |   |
Db     434 TGCTCGGGCCTTTGCAGCTCTGCGCGCTAGTGCACTGGCCCCTCCCGCCGCCGGCCAACA 375

Qy     680 GACGTCTGCACGGGAGGGGGAAGCGGCCCAGGCAGACAGATGGCAGGCGTTCGGACTGGA 739
           |  | |  | || |||  |   ||||| | ||  | | |  | |||  | ||  |  |||
Db     374 GCAGCCNCCGCGCGAGCCGCCGGCGGCTCCGGGCGCCTGGCGCCAGCAGATCCAATGGGA 315

Qy     740 AGATAACGCCGCCGCCTTCAGCCTGGT 766
           | ||||      |  |||||| || |
Db     314 GAACAACGGGCAGGTGTTCAGCTTGCT 288
```

```
RESULT   5
AI006602
LOCUS       AI006602      549 bp    mRNA            EST       12-JUN-1998
DEFINITION  ue15g11.y1 Sugano mouse embryo mewa Mus musculus cDNA clone
            IMAGE:1480484 5' similar to gb:U22399 Mus musculus Cdk-inhibitor
            p57KIP2 (MOUSE);, mRNA sequence.
ACCESSION   AI006602
NID         g3216211
VERSION     AI006602.1  GI:3216211
KEYWORDS    EST.
```

```
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 549)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On May 9, 1995 this sequence version replaced gi:803086.

            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:928840
            Seq primer: primer name ambiguous
            High quality sequence stop: 489.
FEATURES             Location/Qualifiers
     source          1. .549
                     /organism="Mus musculus"
                     /strain="C57BL"
                     /db_xref="taxon:10090"
                     /map="13; 19p13.1"
                     /clone="IMAGE:1480484"
                     /clone_lib="Sugano mouse embryo mewa"
                     /dev_stage="embryo, 14 dpc"
                     /lab_host="DH10B"
                     /note="Vector: pME18S-FL3; Site_1: DraIII (CACTGTGTG);
                     Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed
                     with an oligo(dT) primer  [ATGTGGCCTTTTTTTTTTTTTTTT];
                     double-stranded cDNA was ligated to a  DraIII adaptor
                     [TGTTGGCCTACTGG], digested and cloned into distinct DraIII
                     sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site
                     CACCATGTG). XhoI should be used to isolate the cDNA
                     insert.  Size selection was performed to exclude fragments
                     <1.5kb.  Library constructed by Dr. Sumio Sugano
                     (University of Tokyo Institute of Medical Science).
                     Custom primers  for sequencing: 5' end primer
                     CTTCTGCTCTAAAAGCTGCG and 3' end primer
                     CGACCTGCAGCTCGAGCACA."
BASE COUNT       93 a    183 c    191 g     81 t      1 others
ORIGIN
```

```
  Query Match             2.4%;  Score 37;  DB 41;  Length 549;
  Best Local Similarity   54.5%;  Pred. No. 3.2;
  Matches   73;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy     632 CCGTCCCCGGGAAGCCCGGGCACTGCACCTGGCGGTCGCTGACTGGCTGACGTCTGCACG 691
           ||  |  | ||   ||||| |||  |   |||| || |||| |   | |   ||||  |
Db     416 CCTGCAGCTGGGGCCCCGGCCACCCCCGGTGGCCGTGGCTGTCATCCCGCGTTCTGGGCC 475

Qy     692 GGAGGGGGAAGCGGCCCAGGCAGACAGATGGCAGGCGTTCGGACTGGAAGATAACGCCGC 751
           |  ||  |  | |||||  |  | |  |  | |||||   |  | |       ||||||  |
Db     476 GCCGGCTGGCGAGGCCCCCGACGGCCTAGAGGAGGCGCCTGAGCAGCCGNCCAACGCCCC 535

Qy     752 CGCCTTCAGCCTGG 765
           ||||||  | |||
Db     536 AGCCTTCGCCGTGG 549
```

```
RESULT   6
AI305207/c
LOCUS       AI305207      474 bp    mRNA            EST       29-JAN-1999
```

```
DEFINITION  ql90d04.x1 Soares_NhHMPu_S1 Homo sapiens cDNA clone IMAGE:1879591
            3' similar to gb:S78694 PROTEIN-LYSINE 6-OXIDASE PRECURSOR
            (HUMAN);contains TAR1.t1 TAR1 repetitive element ;, mRNA sequence.
ACCESSION   AI305207
NID         g3988848
VERSION     AI305207.1  GI:3988848
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 474)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Feb 11, 1998 this sequence version replaced gi:2871419.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 744   Std Error: 0.00
            Seq primer: -40UP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .474
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1879591"
                     /clone_lib="Soares_NhHMPu_S1"
                     /tissue_type="Pooled human melanocyte, fetal heart, and
                     pregnant uterus"
                     /lab_host="DH10B"
                     /note="Organ: mixed (see below); Vector: pT7T3D-Pac
                     (Pharmacia) with a modified polylinker; Site_1: Not I;
                     Site_2: Eco RI; Equal amounts of plasmid DNA from three
                     normalized libraries (melanocyte 2NbHM, pregnant uterus
                     NbHPU, and fetal heart NbHH19W) were mixed, and ss circles
                     were made in vitro. Following HAP purification, this DNA
                     was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from pools of
                     5,000 clones made from the same 3 libraries. The pools
                     consisted of I.M.A.G.E. clones 260232-265223,
                     340488-345479, and 484488-489479."
BASE COUNT       62 a    147 c    195 g     70 t
ORIGIN
```

```
  Query Match             2.4%;  Score 36.6;  DB 44;  Length 474;
  Best Local Similarity   53.1%;  Pred. No. 3.9;
  Matches   78;  Conservative    0;  Mismatches   69;  Indels    0;  Gaps    0;

Qy     620 TGGCGGGGCCTTCCGTCCCCGGGAAGCCCGGGCACTGCACCTGGCGGTCGCTGACTGGCT 679
           ||   |||||||     | | |   ||  | |||||||  ||   | | ||| | |   |
Db     452 TGCTCGGGCCTTTGCAGCTCTGCGCGCTAGTGCACTGCCCCCTCCCGCCGCCGGCCAACA 393

Qy     680 GACGTCTGCACGGGAGGGGGAAGCGGCCCAGGCAGACAGATGGCAGGCGTTCGGACTGGA 739
           |  | |    || |||  |   ||||| | ||  | | |  | |||  | ||  |  |||
Db     392 GCAGCCCCGCCGCGAGCCGCCGGCGGCTCCGGGCGCCTGGCGCCAGCAGATCCAATGGGA 333

Qy     740 AGATAACGCCGCCGCCTTCAGCCTGGT 766
           | ||||     |  |||||| || |
Db     332 GAACAACGGGCAGGTGTTCAGCTTGCT 306
```

```
RESULT   7
C69795
LOCUS       C69795        373 bp    mRNA            EST       29-SEP-1997
DEFINITION  C69795 Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
            clone yk366e8 5', mRNA sequence.
ACCESSION   C69795
NID         g2440320
```

```
VERSION     C69795.1  GI:2440320
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 373)
  AUTHORS   Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,
            Sano,M., Miyata,A. and Nishigaki,A.
  TITLE     Expression map of the C.elegans genome
  JOURNAL   Unpublished (1996)
COMMENT
            Contact: Yuji Kohara
            Gene Library Lab
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411, Japan
            Tel: 0559-75-0771
            Fax: 0559-75-6240
            Email: ykohara@ddbj.nig.ac.j.
FEATURES              Location/Qualifiers
     source           1. .373
                      /organism="Caenorhabditis elegans"
                      /strain="CB1489 him-8(e1489)"
                      /db_xref="taxon:6239"
                      /clone="yk366e8"
                      /clone_lib="Yuji Kohara unpublished cDNA"
                      /note="dev_stage=varied, sex=Hermaphrodite male,
                      tissue_type=whole animal"
BASE COUNT       81 a     89 c    127 g     74 t      2 others
ORIGIN


  Query Match             2.4%;  Score 36.2;  DB 36;  Length 373;
  Best Local Similarity   50.0%;  Pred. No. 4.4;
  Matches   89;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy    126 CCTGCCGGAGTTGCCGGAAGGACTACGGGAGCTGGAGGTCTCTGGTAACCTACAACTGAC 185
          ||||| ||||  | |||  | |  | |  |||||||||      |   |||  || |||
Db     88 CCTGCTGGAGGAGGCGGTGGCAACAAGTCAGCTGGAGGAGGTGGTGCACCACCACCTGGA 147

Qy    186 CAGCCTGCCATCGCTGCCGCAGGGACTACAGAAGCTGTGGGCCTATAATAATTGGCTGGC 245
          | |      |   |  || ||  |   || ||||   |   ||   ||| |||| |
Db    148 GTGTCGTGTTACTTGGGAGCTGGAGCCGGAGGCGCTGCCAGTGGATCTCAATCGGCTTAC 207

Qy    246 CAGCCTGCCGACGTTGCCGCCAGGACTAGGGGATCTGGCGGTCTCTAATAACCAGCTG 303
            || ||  |  | | |        | |||  |||||||     |  ||||||||
Db    208 TTTGGTGTCGGAGGAGGCCCAGTTGGAGGTGGAGGTGGCGGTGGANCACCACCAGCTG 265


RESULT   8
C67933
LOCUS       C67933         360 bp    mRNA            EST       22-SEP-1997
DEFINITION  C67933 Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
            clone yk307h10 5', mRNA sequence.
ACCESSION   C67933
NID         g2426863
VERSION     C67933.1  GI:2426863
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 360)
  AUTHORS   Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,
            Sano,M., Miyata,A. and Nishigaki,A.
  TITLE     Expression map of the C.elegans genome
  JOURNAL   Unpublished (1996)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1407255.

            Contact: Yuji Kohara
            Gene Library Lab
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411, Japan
            Tel: 0559-75-0771
            Fax: 0559-75-6240
            Email: ykohara@ddbj.nig.ac.j.
FEATURES              Location/Qualifiers
     source           1. .360
                      /organism="Caenorhabditis elegans"
                      /strain="CB1489 him-8(e1489)"
                      /db_xref="taxon:6239"
                      /clone="yk307h10"
                      /clone_lib="Yuji Kohara unpublished cDNA"
                      /note="dev_stage=varied, sex=Hermaphrodite male,
                      tissue_type=whole animal"
BASE COUNT       72 a     86 c    118 g     74 t      10 others
ORIGIN


  Query Match             2.3%;  Score 35.8;  DB 36;  Length 360;
  Best Local Similarity   49.4%;  Pred. No. 5.7;
  Matches   88;  Conservative    0;  Mismatches   90;  Indels    0;  Gaps    0;

Qy    126 CCTGCCGGAGTTGCCGGAAGGACTACGGGAGCTGGAGGTCTCTGGTAACCTACAACTGAC 185
          ||||| ||||  | |||  | |  | |  |||||||||      |   |||  || |||
Db     71 CCTGCTGGAGGAGGCGGTGGCAACAAGTCAGCTGGAGGAGGTGGTGCACCACCACCTGGA 130

Qy    186 CAGCCTGCCATCGCTGCCGCAGGGACTACAGAAGCTGTGGGCCTATAATAATTGGCTGGC 245
          | |      |   |  || ||  |   || ||||   |   ||   ||| |||| |
Db    131 GTGTCGTGTTACTTGGGAGCTGGAGCCGGAGGCGCTGCCAGTGGATCTCAATCGGCTTAC 190

Qy    246 CAGCCTGCCGACGTTGCCGCCAGGACTAGGGGATCTGGCGGTCTCTAATAACCAGCTG 303
            || ||  |  | | |        | |||  |||||||     |  ||||||||
Db    191 TTTGGTGTCGGAGGANGCCCANTTGGANGTGGAAGTGGCGGTGGAGCACCACCAGCTG 248


RESULT   9
C69422
LOCUS       C69422         376 bp    mRNA            EST       23-SEP-1997
DEFINITION  C69422 Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
            clone yk363d4 5', mRNA sequence.
ACCESSION   C69422
NID         g2430778
VERSION     C69422.1  GI:2430778
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 376)
  AUTHORS   Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,
            Sano,M., Miyata,A. and Nishigaki,A.
  TITLE     Expression map of the C.elegans genome
  JOURNAL   Unpublished (1996)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1294168.

            Contact: Yuji Kohara
            Gene Library Lab
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411, Japan
            Tel: 0559-75-0771
            Fax: 0559-75-6240
            Email: ykohara@ddbj.nig.ac.j.
FEATURES              Location/Qualifiers
     source           1. .376
                      /organism="Caenorhabditis elegans"
                      /strain="CB1489 him-8(e1489)"
                      /db_xref="taxon:6239"
                      /clone="yk363d4"
                      /clone_lib="Yuji Kohara unpublished cDNA"
                      /note="dev_stage=varied, sex=Hermaphrodite male,
                      tissue_type=whole animal"
BASE COUNT       76 a     89 c    127 g     78 t      6 others
ORIGIN
```

```
  Query Match             2.3%;  Score 35.8;  DB 36;  Length 376;
  Best Local Similarity   49.4%;  Pred. No. 5.8;
  Matches   88;  Conservative    0;  Mismatches   90;  Indels    0;  Gaps    0;

Qy    126 CCTGCCGGAGTTGCCGGAAGGACTACGGGAGCTGGAGGTCTCTGGTAACCTACAACTGAC 185
          ||||| |||| | ||| | | | | ||||||||     |   ||| || |||
Db     84 CCTGCTGGAGGAGGCGGTGGCAACAAGTCAGCTGGAGGAGGTGGTGCACCACCACCTGGA 143

Qy    186 CAGCCTGCCATCGCTGCCGCAGGGACTACAGAAGCTGTGGGCCTATAATAATTGGCTGGC 245
          | |       |   | || ||  |   ||  ||||   |   ||   ||| ||||  |
Db    144 GTGTCGTGTTACTTGGGAGCTGGAGCCGGAGGCGCTGCCAGTGGATCTCAATCGGCTTAC 203

Qy    246 CAGCCTGCCGACGTTGCCGCCAGGACTAGGGGATCTGGCGGTCTCTAATAACCAGCTG 303
            || ||  |    | |       | |||  |||||||     |  ||||||||
Db    204 TTTGGTGTCGGAGGANGCCCANTTGGAGGTGGANGTGGCGGTGGAGCACCACCAGCTG 261


RESULT  10
D69334
LOCUS       D69334       360 bp    mRNA            EST       07-DEC-1995
DEFINITION  CELK068F1F Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
            clone yk68f1 5', mRNA sequence.
ACCESSION   D69334
NID         g1104979
VERSION     D69334.1  GI:1104979
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 360)
  AUTHORS   Kohara,Y., Mitsuki,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and
            Tabara,H.
  TITLE     Toward an expression map of the C.elegans genome
  JOURNAL   Unpublished (1994)
COMMENT     On Apr 14, 1993 this sequence version replaced gi:837869.

            Contact: Yuji Kohara
            Gene Library Lab
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411, Japan
            Tel: 0559-75-0771
            Fax: 0559-75-6240
            Email: ykohara@ddbj.nig.ac.j
            High quality sequence stop: 163.
FEATURES             Location/Qualifiers
     source          1. .360
                     /organism="Caenorhabditis elegans"
                     /strain="CB1489 him-8(e1489)"
                     /db_xref="taxon:6239"
                     /clone="yk68f1"
                     /clone_lib="Yuji Kohara unpublished cDNA"
                     /note="dev_stage=varied, sex=Hermaphrodite male,
                     tissue_type=whole animal"
BASE COUNT       75 a     84 c    128 g     71 t      2 others
ORIGIN


  Query Match             2.3%;  Score 35.6;  DB 24;  Length 360;
  Best Local Similarity   50.0%;  Pred. No. 6.5;
  Matches   89;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy    126 CCTGCCGGAGTTGCCGGAAGGACTACGGGAGCTGGAGGTCTCTGGTAACCTACAACTGAC 185
          ||||| |||| | ||| | | | | ||||||||     |   ||| || |||
Db     88 CCTGCTGGAGGAGGCGGTGGCAACAAGTCAGCTGGAGGAGGTGGTGCACCACCACCTGGA 147

Qy    186 CAGCCTGCCATCGCTGCCGCAGGGACTACAGAAGCTGTGGGCCTATAATAATTGGCTGGC 245
          | |       |   | || ||  |   ||  ||||   |   ||   ||| ||||  |
Db    148 GTGTCGTGTTACTTGGGAGCTGGAGCCGGAGGCGCTGCCAGTGGATCTCAATCGGCTTAC 207

Qy    246 CAGCCTGCCGACGTTGCCGCCAGGACTAGGGGATCTGGCGGTCTCTAATAACCAGCTG 303
            || ||  | | | |       | |||  |||||||     |  ||||||||
Db    208 TTTGGTGTCGGAGGAGGCCCAGTTGGAGGTGGAGGTGGCGGTGGAGCACCACCAGCTG 265


RESULT  11
AI402251/c
LOCUS       AI402251      647 bp    mRNA            EST       08-FEB-1999
DEFINITION  GH11218.3prime GH Drosophila melanogaster head pOT2 Drosophila
            melanogaster cDNA clone GH11218 3prime, mRNA sequence.
ACCESSION   AI402251
NID         g4245338
VERSION     AI402251.1  GI:4245338
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 647)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     On Apr 7, 1998 this sequence version replaced gi:3036223.
            Other_ESTs: GH11218.5prime
            Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Based upon the presence of a XhoI site followed by a run of 14 or
            more T residues at the beginning of the sequence, this clone was
            polyadenylated. The resulting Poly-T sequence has been removed.
            Plate: 112  row: B  column: 6
            High quality sequence stop: 637.
FEATURES             Location/Qualifiers
     source          1. .647
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /map="15"
                     /clone="GH11218"
                     /clone_lib="GH Drosophila melanogaster head pOT2"
                     /sex="male and female"
                     /dev_stage="adult"
                     /lab_host="DH5 - alpha"
                     /note="Organ: head; Vector: pOT2; Site_1: EcoRI; Site_2:
                     XhoI; Sized fractionated cDNAs were directly ligated into
                     pOT2.  Plasmid cDNA library."
BASE COUNT      127 a    172 c    170 g    178 t
ORIGIN


  Query Match             2.3%;  Score 35.6;  DB 46;  Length 647;
  Best Local Similarity   54.6%;  Pred. No. 8.9;
  Matches   71;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy     90 TACAACACTGGTTATTCCTGATAATAATCTGACCAGCCTGCCGGAGTTGCCGGAAGGACT 149
          |||||    |||   ||| || || |    || | | |||| | ||| |  | |||
Db    551 TACAAGGAAGGTGTCTCCCGAGAACACAAAGAACCGTCTGCAGAAGTGGGAGCAAGCCAA 492

Qy    150 ACGGGAGCTGGAGGTCTCTGGTAACCTACAACTGACCAGCCTGCCATCGCTGCCGCAGGG 209
          |||||||  | ||       |  || | ||  |||| | | ||||  |       ||||
Db    491 GCGGGAGCAGCAGCAACAGGAAAAACAACGCCTGATCCGACTGCAGTGCAACAAGCAGAA 432

Qy    210 ACTACAGAAG 219
          |||  || ||
Db    431 ACTGGAGGAG 422


RESULT  12
AI660752
LOCUS       AI660752      547 bp    mRNA            EST       10-MAY-1999
DEFINITION  we66h08.x1 Soares_thymus_NHFTh Homo sapiens cDNA clone
            IMAGE:2346111 3' similar to TR:O75300 O75300 ES/130. ;, mRNA
```

```
            sequence.
ACCESSION   AI660752
NID         g4764335
VERSION     AI660752.1  GI:4764335
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 547)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Mar 20, 1998 this sequence version replaced gi:2979826.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40UP from Gibco
            High quality sequence stop: 443.
FEATURES             Location/Qualifiers
     source          1. .547
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2346111"
                     /clone_lib="Soares_thymus_NHFTh"
                     /dev_stage="fetal"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: thymus, pooled; Vector: pT7T3D-Pac
                     (Pharmacia) with a modified polylinker; Site_1: Not I;
                     Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                     oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCAACGTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector.  Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M. Fatima Bonaldo.  "
BASE COUNT      158 a    142 c    177 g     69 t      1 others
ORIGIN
```

```
  Query Match             2.3%;  Score 35.4;  DB 49;  Length 547;
  Best Local Similarity   55.9%;  Pred. No. 9.3;
  Matches   66;  Conservative    0;  Mismatches   52;  Indels    0;  Gaps    0;

Qy     953 ACAACTGGTACATAATGCAGAAAAAGGGGAATACGACAACAATCTCCAGGGGCTGGTTTC 1012
           ||| |||    | |    |||||||| |||| | |  |  | |  || |||||  |   |
Db     391 ACAGCTGAAGGAGAGACTAGAAAAAGAGGAAGAAGTAACAAGTGACCTGGGGCGCGCCGC 450

Qy    1013 CACGGGGCGTGAGATGTTCCGCCTGGCAACACTGGAACAGATTGCCCGGGAAAAAGCC 1070
           |||| | |   ||  | | |    | | || |  | ||||  | |||  || | ||| |
Db     451 CACGAGACTGCAGGAGCTTCTGAAGACGACCCANGGACAGCTGGCCAAGGGAGAAGAC 508
```

```
RESULT  13
D61630
LOCUS       D61630         353 bp    mRNA            EST       14-DEC-1995
DEFINITION  HUM424H04B Clontech human fetal brain polyA+ mRNA (#6535) Homo
            sapiens cDNA clone GEN-424H04 5', mRNA sequence.
ACCESSION   D61630
NID         g970903
VERSION     D61630.1  GI:970903
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 353)
  AUTHORS   Fujiwara,T., Hirano,H., Katagiri,T., Kawai,A., Kuga,Y., Nagata,M.,
            Okuno,S., Ozaki,K., Shimizu,F., Shimada,Y., Shinomiya,H.,
            Takaichi,A., Takeda,S., Watanabe,T., Takahashi,E., Hirai,Y.,
            Maekawa,H., Shin,S. and Nakamura,Y.
  TITLE     Fujiwara et al. (1995)
  JOURNAL   Unpublished (1995)
COMMENT     On Apr 14, 1993 this sequence version replaced gi:785764.

            Contact: Tsutomu Fujiwara
            Otsuka GEN Research Institute
            Otsuka Pharmaceutical Co.,Ltd
            463-10 kagasuno Kawauchi-cho, Tokushima, Tokushima, 771-01 Japan
            Tel: 0886-65-2888
            Fax: 0886-37-1035
            Insert Length: 1022   Std Error: 0.00
            High quality sequence stop: 262.
FEATURES             Location/Qualifiers
     source          1. .353
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="3 p24.2-p21.1"
                     /clone="GEN-424H04"
                     /clone_lib="Clontech human fetal brain polyA+ mRNA
                     (#6535)"
BASE COUNT      58 a    105 c    123 g     61 t      6 others
ORIGIN
```

```
  Query Match             2.3%;  Score 35.2;  DB 24;  Length 353;
  Best Local Similarity   53.0%;  Pred. No. 8.3;
  Matches   70;  Conservative    2;  Mismatches   60;  Indels    0;  Gaps    0;

Qy     620 TGGCGGGGCCTTCCGTCCCCGGGAAGCCCGGGCACTGCACCTGGCGGTCGCTGACTGGCT 679
           |||  ||| || | || |  || |  :||   |    |     |||:||  |||| ||
Db     221 TGGGCGGGGCTGCTGTGCTGGGCATTSCCACCCTGGCCGTGAAGCGKTCATTGACAGGGC 280

Qy     680 GACGTCTGCACGGGAGGGGGAAGCGGCCCAGGCAGACAGATGGCAGGCGTTCGGACTGGA 739
            ||     | ||||| | ||| |   || ||||||||||| |||  | | |  |  ||  |
Db     281 CACTAGCCCGCGGGATGAGGATGACACCAAGGCAGACAGCTGGAGGACCTAGGCTCTCAA 340

Qy     740 AGATAACGCCGC 751
           |   ||  | |
Db     341 GGCACACCACAC 352
```

```
RESULT  14
AU045269
LOCUS       AU045269       403 bp    mRNA            EST       09-DEC-1998
DEFINITION  AU045269 Mouse sixteen-cell-embryo cDNA Mus musculus cDNA clone
            J0932D09 3', mRNA sequence.
ACCESSION   AU045269
NID         g3981452
VERSION     AU045269.1  GI:3981452
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 403)
  AUTHORS   Ko,M.S.H., Kitchen,J.R., Wang,X., Threat,T.A., Sun,T.,
            DePalma,G.E., Liang,Y., Kargul,G.J., Sharara,R., Lim,M.K. and
            Doi,H.
  TITLE     Systematic analyses of genes expressed in 16-cell mouse embryo (The
            ERATO/Doi Project at Wayne State University)
  JOURNAL   Unpublished (1998)
COMMENT     On May 8, 1995 this sequence version replaced gi:801081.

            Contact: Hirofumi Doi
            Doi Bioasymmetry Project, ERATO
            Japan Science and Technology Corporation (JST)
            WBG Marive East 12F, 2-6 Nakase, Mihama-ku, Chiba 261-71, Japan
            Email: hd@bioa.jst.go.jp.
FEATURES             Location/Qualifiers
     source          1. .403
```

```
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="J0932D09"
                     /clone_lib="Mouse sixteen-cell-embryo cDNA"
                     /dev_stage="sixteen-cell-embryo"
BASE COUNT       130 a     71 c     91 g    105 t      6 others
ORIGIN


  Query Match             2.3%;  Score 35.2;  DB 44;  Length 403;
  Best Local Similarity   50.6%;  Pred. No. 8.9;
  Matches   85;  Conservative    0;  Mismatches   83;  Indels    0;  Gaps    0;

Qy     916 ACCGGGTCACACATGCCCTGCACCAGATGAATAACGTACAACTGGTACATAATGCAGAAA 975
           ||   | || |    |||    ||||   |||| ||||| |||  || ||||    |
Db     132 ACAATGACAGAAGGTCCCATTTCCAGTATTATAAGGTACATGTGGCCCAAAATGGCTGAG 191

Qy     976 AAGGGGAATACGACAACAATCTCCAGGGGCTGGTTTCCACGGGGCGTGAGATGTTCCGCC 1035
           ||     |||  |||| ||   | |   || |     |  ||||| ||   ||   | ||
Db     192 AATAAACATAAAACAAGAAAGGCAAAAAGCAGTAAAGCTAGGGGCATGCATTGGGGCTCC 251

Qy    1036 TGGCAACACTGGAACAGATTGCCCGGGAAAAAGCCGGAACACTGGCTT 1083
            ||  | | |   | | ||  |||       | | ||| || ||| ||
Db     252 AGGAGAAAATCTCAGAAATACCCCAATGTGCAACGGGAGCAATGGGTT 299


RESULT  15
AI287656/c
LOCUS       AI287656      494 bp    mRNA            EST       24-NOV-1998
DEFINITION  ql53a03.x1 Soares_NhHMPu_S1 Homo sapiens cDNA clone IMAGE:1876012
            3' similar to gb:S78694 PROTEIN-LYSINE 6-OXIDASE PRECURSOR
            (HUMAN);contains TAR1.t3 TAR1 repetitive element ;, mRNA sequence.
ACCESSION   AI287656
NID         g3927409
VERSION     AI287656.1  GI:3927409
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 494)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2151894.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40UP from Gibco
            High quality sequence stop: 492.
FEATURES             Location/Qualifiers
     source          1. .494
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="633E01; 17; 17p11.2-p13; 17p12-17p13"
                     /clone="IMAGE:1876012"
                     /clone_lib="Soares_NhHMPu_S1"
                     /tissue_type="Pooled human melanocyte, fetal heart, and
                     pregnant uterus"
                     /lab_host="DH10B"
                     /note="Organ: mixed (see below); Vector: pT7T3D-Pac
                     (Pharmacia) with a modified polylinker; Site_1: Not I;
                     Site_2: Eco RI; Equal amounts of plasmid DNA from three
                     normalized libraries (melanocyte 2NbHM, pregnant uterus
                     NbHPU, and fetal heart NbHH19W) were mixed, and ss circles
                     were made in vitro. Following HAP purification, this DNA
                     was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from pools of
                     5,000 clones made from the same 3 libraries. The pools
                     consisted of I.M.A.G.E. clones 260232-265223,
                     340488-345479, and 484488-489479."
BASE COUNT       65 a    153 c    199 g     77 t
ORIGIN


  Query Match             2.3%;  Score 35;  DB 44;  Length 494;
  Best Local Similarity   52.4%;  Pred. No. 11;
  Matches   77;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy     620 TGGCGGGGCCTTCCGTCCCCGGGAAGCCCGGGCACTGCACCTGGCGGTCGCTGACTGGCT 679
           ||  |||||||     | | |   ||  | |||||   ||   | | ||| | |   |
Db     452 TGCTCGGGCCTTTGCAGCTCTGCGCGCTAGTGCACTCGCCCCTCCCGCCGCCGGCCAACA 393

Qy     680 GACGTCTGCACGGGAGGGGGAAGCGGCCCAGGCAGACAGATGGCAGGCGTTCGGACTGGA 739
           |  | | |  | || |||  |   ||||| | ||  | | |  | |||  | ||  |  |||
Db     392 GCAGCCCCCGCGCGAGCCGCCGGCGGCTCCGGGCGCCTGGCGCCAGCAGATCCAATGGGA 333

Qy     740 AGATAACGCCGCCGCCTTCAGCCTGGT 766
             | ||||     |  |||||| || |
Db     332 GAACAACGGGCAGGTGTTCAGCTTGCT 306


Search completed: October  5, 1999, 17:14:12
Job time: 8695 sec
```

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 1999, 18:01:16 ; Search time 140.41 Seconds
(without alignments)
1139.653 Million cell updates/sec

Title: US-09-068-804-13
Perfect score: 1538
Sequence: 1 ATGCGTGATTGCCTGAATAA.....CCCGGTGGCGTCTTAACTGA 1538

Scoring table: IDENTITY_NUC

Searched: 192659 seqs, 52021692 residues

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/5C_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/5D_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	45.4	3.0	7218	2	US-08-232-463-14	Sequence 14, Appl
c 2	37.2	2.4	1150	1	US-08-264-861A-10	Sequence 10, Appl
c 3	37.2	2.4	1150	5	PCT-US95-07784-10	Sequence 10, Appl
c 4	35	2.3	30001	1	US-08-125-468-1	Sequence 1, Appli
c 5	35	2.3	30001	3	US-08-474-933-1	Sequence 1, Appli
6	32.4	2.1	2380	3	US-08-572-951-3	Sequence 3, Appli
7	31.8	2.1	31571	1	US-08-323-443B-1	Sequence 1, Appli
c 8	31.8	2.1	16885	2	US-08-390-878-16	Sequence 16, Appl
c 9	31.6	2.1	977	3	US-08-855-140-2	Sequence 2, Appli
10	31.2	2.0	2338	2	US-08-425-069-1	Sequence 1, Appli
11	31	2.0	1995	2	US-08-425-069-3	Sequence 3, Appli
12	30.6	2.0	1393	2	US-08-464-148-3	Sequence 3, Appli
13	30.6	2.0	1393	2	US-08-385-500-3	Sequence 3, Appli
14	30.6	2.0	1393	2	US-08-846-784-3	Sequence 3, Appli
c 15	30.6	2.0	846	5	PCT-US96-05320A-362	Sequence 362, App
16	30.4	2.0	906	1	US-07-817-852A-1	Sequence 1, Appli
17	30.4	2.0	906	3	US-08-470-137-1	Sequence 1, Appli
18	30.2	2.0	801	1	US-07-959-946-4	Sequence 4, Appli
19	30.2	2.0	3725	1	US-08-155-331-12	Sequence 12, Appl
20	30.2	2.0	349	1	US-08-275-225-3	Sequence 3, Appli
21	30.2	2.0	3725	2	US-08-424-022-12	Sequence 12, Appl
22	30.2	2.0	964	2	US-08-448-606-5	Sequence 5, Appli
23	30.2	2.0	863	2	US-08-448-606-7	Sequence 7, Appli
24	30.2	2.0	349	2	US-08-644-729-18	Sequence 18, Appl
25	30.2	2.0	801	3	US-08-333-577-4	Sequence 4, Appli
26	30.2	2.0	3262	3	US-08-678-039A-41	Sequence 41, Appl
27	30.2	2.0	3725	4	US-08-424-017B-12	Sequence 12, Appl
28	30.2	2.0	801	5	PCT-US92-08634-4	Sequence 4, Appli
29	30.2	2.0	3725	5	PCT-US93-11696-12	Sequence 12, Appl
c 30	30	2.0	10898	4	US-08-481-658B-5	Sequence 5, Appli
c 31	30	2.0	1174	4	US-08-481-658B-39	Sequence 39, Appl
c 32	30	2.0	1401	4	US-08-481-658B-49	Sequence 49, Appl
33	29.8	1.9	2647	1	US-08-136-119-3	Sequence 3, Appli
34	29.6	1.9	349	1	US-08-275-225-10	Sequence 10, Appl
35	29.6	1.9	439	2	US-08-644-729-27	Sequence 27, Appl
c 36	29.6	1.9	1302	3	US-08-529-600D-3	Sequence 3, Appli
c 37	29.6	1.9	1302	4	US-08-973-275-6	Sequence 6, Appli
c 38	29.4	1.9	2277	3	US-08-496-855A-1	Sequence 1, Appli
c 39	29.4	1.9	2068	3	US-08-466-589-1	Sequence 1, Appli
c 40	29.4	1.9	1425	3	US-08-578-551-2	Sequence 2, Appli
c 41	29.4	1.9	2068	4	US-08-700-636-1	Sequence 1, Appli
c 42	29.4	1.9	12588	4	US-08-387-942C-1	Sequence 1, Appli
c 43	29.2	1.9	2775	2	US-08-481-130-25	Sequence 25, Appl
c 44	29.2	1.9	2927	2	US-08-481-130-27	Sequence 27, Appl
c 45	29.2	1.9	2775	2	US-08-656-984A-25	Sequence 25, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
US-08-232-463-14

Query Match 3.0%; Score 45.4; DB 2; Length 7218;
Best Local Similarity 8.4%; Pred. No. 0.00031;
Matches 37; Conservative 210; Mismatches 196; Indels 0; Gaps 0;

```
Qy    552 TCTGTACGCACTCGTGACAGGCTCTGCGGACATCATTGGCCATTCAGGCATCAGGATACA 611
          ||| | |    ||   | | |   ||  | ||||| ||    | |     | | | |
Db   1520 TCTATTCAGTTTCAAAAAACGGCATGTAGGCATCACTGTAATTACCTATCTATGCAAGTA 1461

Qy    612 CTTCGATATGGCGGGGCCTTCCGTCCCCGGGAAGCCCGGGCACTGCACCTGGCGGTCGCT 671
            ||  | |    |  |  ||  || |  ::::::   ::: :  : :   :: ::  :
```

```
Db   1460 GTTAAAGAGATAGAAGAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1401

Qy    672 GACTGGCTGACGTCTGCACGGGAGGGGGAAGCGGCCCAGGCAGACAGATGGCAGGCGTTC 731
          ::  ::  :: :   : : :::::::::::: ::   ::: ::: ::: :: ::: :
Db   1400 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1341

Qy    732 GGACTGGAAGATAACGCCGCCGCCTTCAGCCTGGTCCTGGACAGACTGCGTGAGACGGAA 791
          :::  :::::: :: :  :  :     ::   ::    ::: :::  : : :::: ::::
Db   1340 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1281

Qy    792 AACTTCAAAAAAGACGCGGGCTTTAAGGCACAGATATCATCCTGGCTGACACAACTGGCT 851
          ::    :::::::: : :::    :::: : ::: :  :    ::  :: : ::  ::
Db   1280 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1221

Qy    852 GAAGATGCTGCGCTGAGAGCAAAAACCTTTGCCATGGCAACAGAGGCAACATCAACCTGC 911
          ::::: :  : :  ::::: :::::     :  : :: :: ::::: :: :  ::   :
Db   1220 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1161

Qy    912 GAGGACCGGGTCACACATGCCCTGCACCAGATGAATAACGTACAACTGGTACATAATGCA 971
          :::::  :::  : : : :    : :  ::: ::: :: : : ::  :: : : :: : :
Db   1160 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1101

Qy    972 GAAAAAGGGGAATACGACAACAA 994
          :::::::::::: : :: :: ::
Db   1100 RRRRRRRRRRRRRRRRRRRRRRR 1078
```

```
RESULT   2
US-08-264-861A-10/c
; Sequence 10, Application US/08264861A
; Patent No. 5622866
;  GENERAL INFORMATION:
;    APPLICANT:  MOTAMEDI, HAIDEH
;    APPLICANT:  SHAFIEE, ALI
;    TITLE OF INVENTION:  EXPRESSION CASSETTES USEFUL IN
;    TITLE OF INVENTION:  CONSTRUCTION OF INTEGRATIVE AND REPLICATIVE EXPRESSION
;    TITLE OF INVENTION:  VECTORS FOR STREPTOMYCES
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  J. ERIC THIES
;      STREET:  P.O. BOX 2000, 126 E. LINCOLN AVENUE
;      CITY:  RAHWAY
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  07065
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/264,861A
;      FILING DATE:  23-JUN-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  THIES, J. ERIC
;      REGISTRATION NUMBER:  35,382
;      REFERENCE/DOCKET NUMBER:  19132
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (908)594-3904
;      TELEFAX:  (908) 594-4720
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1150 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-264-861A-10
```

```
  Query Match             2.4%;  Score 37.2;  DB 1;  Length 1150;
  Best Local Similarity   52.6%;  Pred. No. 0.04;
  Matches   81;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy    622 GCGGGGCCTTCCGTCCCCGGGAAGCCCGGGCACTGCACCTGGCGGTCGCTGACTGGCTGA 681
          |||| ||| || | |  |||   | || || ||   ||   | ||  || ||  |  | |
Db   1130 GCGGAGCCCTCGGACATCGGCGGGGCCTGGGACGCGACGCTGGGGCAGCGGATCGCGTCA 1071

Qy    682 CGTCTGCACGGGAGGGGGAAGCGGCCCAGGCAGACAGATGGCAGGCGTTCGGACTGGAAG 741
          |   | |   | || |  |  | || || | | |  | | ||||| || || || ||
Db   1070 CTGACGGAGTCGGAGCGGCGGGTGGCCGCGCTGGCCGCGGTCGGGCGTACGAACAGGGAG 1011

Qy    742 ATAACGCCGCCGCCTTCAGCCTGGTCCTGGACAG 775
          ||  |   || ||  |  | |  | || | || |
Db   1010 ATCGCCGAGCAGCTGTTCGTCACGGCCAGCACGG 977
```

```
RESULT   3
PCT-US95-07784-10/c
; Sequence 10, Application PC/TUS9507784
;  GENERAL INFORMATION:
;    APPLICANT:  MOTAMEDI, HAIDEH
;    APPLICANT:  SHAFIEE, ALI
;    TITLE OF INVENTION:  EXPRESSION CASSETTES USEFUL IN
;    TITLE OF INVENTION:  CONSTRUCTION OF INTEGRATIVE AND REPLICATIVE EXPRESSION
;    TITLE OF INVENTION:  VECTORS FOR STREPTOMYCES
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  J. ERIC THIES
;      STREET:  P.O. BOX 2000, 126 E. LINCOLN AVENUE
;      CITY:  RAHWAY
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  07065
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/07784
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  THIES, J. ERIC
;      REGISTRATION NUMBER:  35,382
;      REFERENCE/DOCKET NUMBER:  19132
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (908)594-3904
;      TELEFAX:  (908) 594-4720
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1150 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
PCT-US95-07784-10
```

```
  Query Match             2.4%;  Score 37.2;  DB 5;  Length 1150;
  Best Local Similarity   52.6%;  Pred. No. 0.04;
  Matches   81;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy    622 GCGGGGCCTTCCGTCCCCGGGAAGCCCGGGCACTGCACCTGGCGGTCGCTGACTGGCTGA 681
          |||| ||| || | |  |||   | || || ||   ||   | ||  || ||  |  | |
Db   1130 GCGGAGCCCTCGGACATCGGCGGGGCCTGGGACGCGACGCTGGGGCAGCGGATCGCGTCA 1071

Qy    682 CGTCTGCACGGGAGGGGGAAGCGGCCCAGGCAGACAGATGGCAGGCGTTCGGACTGGAAG 741
          |   | |   | | |  |  | || || | | |  | | ||||| || || || ||
Db   1070 CTGACGGAGTCGGAGCGGCGGGTGGCCGCGCTGGCCGCGGTCGGGCGTACGAACAGGGAG 1011
```

```
Qy    742 ATAACGCCGCCGCCTTCAGCCTGGTCCTGGACAG 775
          ||  |   || ||  |  | |  | || | || |
Db   1010 ATCGCCGAGCAGCTGTTCGTCACGGCCAGCACGG 977


RESULT   4
US-08-125-468-1/c
; Sequence 1, Application US/08125468
; Patent No. 5589385
;  GENERAL INFORMATION:
;    APPLICANT:  Ryan, Michael J.
;    APPLICANT:  Lotvin, Jason A.
;    APPLICANT:  Strathy, Nancy
;    APPLICANT:  Fantini, Susan E.
;    TITLE OF INVENTION:  Cloning of the biosynthetic pathway for
;    TITLE OF INVENTION:  chlortetracycline and tetracyline Formation and cosmids
;    TITLE OF INVENTION:  useful therein
;    NUMBER OF SEQUENCES:  1
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  American Cyanamid Company
;      STREET:  One Cyanamid Plaza
;      CITY:  Wayne
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07470
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/125,468
;      FILING DATE:  22-SEP-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Tsevdos, Estelle J
;      REGISTRATION NUMBER:  31,145
;      REFERENCE/DOCKET NUMBER:  31,255-02
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (201)831-3241
;      TELEFAX:  (201)831-3305
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  30001 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-125-468-1


  Query Match             2.3%;  Score 35;  DB 1;  Length 30001;
  Best Local Similarity   52.4%;  Pred. No. 1.4;
  Matches   77;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy    621 GGCGGGGCCTTCCGTCCCCGGGAAGCCCGGGCACTGCACCTGGCGGTCGCTGACTGGCTG 680
          | | ||||   ||| |    ||  ||   |   | || ||  |||||| | ||| |||
Db  16130 GCCCGGGCGCACCGCCTTGTGGGCGCGGCGCAGCAGCTCCCTGCGGTCCCCGACCGGCCC 16071

Qy    681 ACGTCTGCACGGGAGGGGGAAGCGGCCCAGGCAGACAGATGGCAGGCGTTCGGACTGGAA 740
          |   | ||| | | |   ||||    ||| |    |    ||||||     ||    |||
Db  16070 AGTCCGGCAGGCGTGACCGAAGATCACCACGTCCGCCTCCGGCAGGTCGGTGGCGAAGAA 16011

Qy    741 GATAACGCCGCCGCCTTCAGCCTGGTC 767
          |  | |||||  |   ||  || ||||
Db  16010 GTCACCGCCGTGGAAGTCCACCCGGTC 15984


RESULT   5
US-08-474-933-1/c
; Sequence 1, Application US/08474933
; Patent No. 5866410
;  GENERAL INFORMATION:
;    APPLICANT:  Ryan, Michael J.
;    APPLICANT:  Lotvin, Jason A.
;    APPLICANT:  Strathy, Nancy
;    APPLICANT:  Fantini, Susan E.
;    TITLE OF INVENTION:  Cloning of the biosynthetic pathway for
;    TITLE OF INVENTION:  chlortetracycline and tetracyline Formation and cosmids
;    TITLE OF INVENTION:  useful therein
;    NUMBER OF SEQUENCES:  1
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  American Cyanamid Company
;      STREET:  One Cyanamid Plaza
;      CITY:  Wayne
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07470
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/474,933
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/125,468
;      FILING DATE:  22-SEP-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Tsevdos, Estelle J
;      REGISTRATION NUMBER:  31,145
;      REFERENCE/DOCKET NUMBER:  31,255-02
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (201)831-3241
;      TELEFAX:  (201)831-3305
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  30001 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-474-933-1


  Query Match             2.3%;  Score 35;  DB 3;  Length 30001;
  Best Local Similarity   52.4%;  Pred. No. 1.4;
  Matches   77;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy    621 GGCGGGGCCTTCCGTCCCCGGGAAGCCCGGGCACTGCACCTGGCGGTCGCTGACTGGCTG 680
          | | ||||   ||| |    ||  ||   |   | || ||  |||||| | ||| |||
Db  16130 GCCCGGGCGCACCGCCTTGTGGGCGCGGCGCAGCAGCTCCCTGCGGTCCCCGACCGGCCC 16071

Qy    681 ACGTCTGCACGGGAGGGGGAAGCGGCCCAGGCAGACAGATGGCAGGCGTTCGGACTGGAA 740
          -|   | ||| | | |   ||||    ||| |    |    ||||||     ||    |||
Db  16070 AGTCCGGCAGGCGTGACCGAAGATCACCACGTCCGCCTCCGGCAGGTCGGTGGCGAAGAA 16011

Qy    741 GATAACGCCGCCGCCTTCAGCCTGGTC 767
          |  | |||||  |   ||  || ||||
Db  16010 GTCACCGCCGTGGAAGTCCACCCGGTC 15984


RESULT   6
US-08-572-951-3
; Sequence 3, Application US/08572951
; Patent No. 5824790
;  GENERAL INFORMATION:
;    APPLICANT:  KEELING, PETER L.
;    APPLICANT:  KNIGHT, MARY E.
;    APPLICANT:  GUAN, HANPING
;    TITLE OF INVENTION:  MODIFICATION OF STARCH
;    TITLE OF INVENTION:  SYNTHESIS IN PLANTS
;    NUMBER OF SEQUENCES:  41
```

```
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      ADDRESSEE:  Intellectual Property Group of
;      ADDRESSEE:  Pillsbury Madison & Sutro LLP
;      STREET:  1100 New York Avenue, N.W.
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3918
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/572,951
;      FILING DATE:  15-DEC-1995
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/346,602
;      FILING DATE:  29-NOV-1994
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/263,921
;      FILING DATE:  21-JUN-1994
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Paul N. Kokulis
;      REGISTRATION NUMBER:  16,773
;      REFERENCE/DOCKET NUMBER:  222957/1.02.15C
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 861-3000
;      TELEFAX:  (202) 822-0944
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2380 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-572-951-3


  Query Match              2.1%;  Score 32.4;  DB 3;  Length 2380;
  Best Local Similarity   28.5%;  Pred. No. 2;
  Matches   59;  Conservative   31;  Mismatches  117;  Indels    0;  Gaps    0;

Qy    454 AGGAGCTGGATGTCTCTCATAACCAGCTGACCCGCCTGCCGCAAAGCCTCACGGGTCTGT 513
          | |   | |  |  :  |  ||    :|  | :| :     || :|  | | ||   |
Db    737 AYGTNGTNGGNGCNYTNCCNAARGCNYTNGCNMGNMGNGGNCAYMGNGTNATGGTNGTNA 796

Qy    514 CTTCAGCGGCACGCGTATATCTGGACGGGAATCCACTGTCTGTACGCACTCGTGACAGGC 573
          : |        | | :||:   || |  :     :|    || :| :  :|  |:| :
Db    797 THCCNMGNTAYGGNGARTAYGCNGARGCNMGNGAYYTNGGNGTNMGNMGNMGNTAYAARG 856

Qy    574 TCTGCGGACATCATTGGCCATTCAGGCATCAGGATACACTTCGATATGGCGGGGCCTTCC 633
          |   || ||  |:::   : | |   |:    |:::   :  : |:|| |  | :||:
Db    857 TNGCNGGNCARGAYWSNGARGTNACNTAYTTYCAYWSNTAYATHGAYGGNGTNGAYTTYG 916

Qy    634 GTCCCCGGGAAGCCCGGGCACTGCACC 660
          :   ||:|| |   |  | :  |
Db    917 TNTTYGTNGARGCNCCNCCNTTYMGNC 943


RESULT   7
US-08-323-443B-1
; Sequence 1, Application US/08323443B
; Patent No. 5654170
;  GENERAL INFORMATION:
;    APPLICANT:  KLINGER, KATHERINE W.
;    APPLICANT:  LANDES, GREGORY M.
;    APPLICANT:  BURN, TIMOTHY C.
;    APPLICANT:  CONNORS, TIMOTHY D.
;    APPLICANT:  DACKOWSKI, WILLIAM R.
;    APPLICANT:  GERMINO, GREGORY
;    APPLICANT:  QIAN, FENG
;    TITLE OF INVENTION:  POLYCYSTIC KIDNEY DISEASE GENE
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Darby & Darby PC
;      STREET:  805 Third Avenue
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/323,443B
;      FILING DATE:  12-OCT-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ludwig, S. Peter
;      REGISTRATION NUMBER:  25,351
;      REFERENCE/DOCKET NUMBER:  0372/0A462
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 527-7700
;      TELEFAX:  (212) 753-6237
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  31571 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      CLONE:  PKD1 GENOMIC
US-08-323-443B-1


  Query Match              2.1%;  Score 31.8;  DB 1;  Length 31571;
  Best Local Similarity   56.1%;  Pred. No. 15;
  Matches   60;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy    558 CGCACTCGTGACAGGCTCTGCGGACATCATTGGCCATTCAGGCATCAGGATACACTTCGA 617
          | |||  | | |  || | | | | ||  ||   ||  ||||   ||    | | |
Db  14877 CACACCTGAGTCCTGCCCAGGGCAGATGCTTCCACACCCAGGGGGCACCTGAGTCCTACC 14936

Qy    618 TATGGCGGGGCCTTCCGTCCCCGGGAAGCCCGGGCACTGCACCTGGC 664
          | ||| |   ||||   ||| ||  ||  ||| ||||||||||||
Db  14937 CAGGGCAGACGCTTCCACACCCTGGGGGCTGGGGGACTGCACCTGGC 14983


RESULT   8
US-08-390-878-16/c
; Sequence 16, Application US/08390878
; Patent No. 5700683
;  GENERAL INFORMATION:
;    APPLICANT:  Stover, Charles K.
;    APPLICANT:  Mahairas, Gregory G.
;    TITLE OF INVENTION:  VIRULENCE-ATTENUAITNG GENETIC DELETIONS
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend Khourie and Crew
;      STREET:  One Market Plaza, Steuart Street Tower, 20th
;      STREET:  Floor
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94105
```

```
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/390,878
;      FILING DATE:  17-FEB-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hunter, Tom
;      REGISTRATION NUMBER:  38,498
;      REFERENCE/DOCKET NUMBER:  15371A-17
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/543/9600
;      TELEFAX:  415/543/5043
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  16885 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-390-878-16


  Query Match             2.1%;  Score 31.8;  DB 2;  Length 16885;
  Best Local Similarity   51.8%;  Pred. No. 10;
  Matches   72;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy   1394 CGGGCTGGTCGACGATACCGATGCAGAACGTACTATCGGAGTAAGTGCGATGGATAGTGC 1453
          | | ||||||     |||| |||| |  |      | ||||    | | |    |||| |||
Db   3580 CTGCCTGGTTCCACATACGGATGAAATAATCCATCTCGGTCAACGCGATCGGGATCGTGT 3521

Qy   1454 GAAAAAAGAATTTCTGGATGGCCTGCGCGCTCTTGTGGATGAGGTGCTGGGTAGCTATCT 1513
            | |  |||    | | |||||     |      ||| ||| ||| | ||  || ||||
Db   3520 TGATACCGAAGAAGTTGGTGGCCGTAAGGACGGCCTGGGTGATGTGGTTGGCGGCGATCT 3461

Qy   1514 GACAGCCCGGTGGCGTCTT 1532
                 ||  |||||| |
Db   3460 CCGGCAGCGACGGCGTCGT 3442


RESULT  9
US-08-855-140-2/c
; Sequence 2, Application US/08855140
; Patent No. 5854022
;  GENERAL INFORMATION:
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Goli, Surya K.
;    APPLICANT:  Guegler, Karl J.
;    TITLE OF INVENTION:  NEW TRANSMEMBRANE PROTEIN
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/855,140
;      FILING DATE:  Herewith
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0296 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  977 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  MYOMNOT01
;      CLONE:  779308
US-08-855-140-2


  Query Match             2.1%;  Score 31.6;  DB 3;  Length 977;
  Best Local Similarity   51.7%;  Pred. No. 2.1;
  Matches   61;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy   1235 TGAAAACAGCGGGTTCAGTAAATGGATACTGCAGTGGGGGCCGTTACACAGCGTGCTGGA 1294
          ||||  ||||     |||| | ||| ||       | ||||  | | ||  |  |  ||||
Db    974 TGAANCCAGCCTAANCAGTTANTGGGTAGGCACTTTGGGGTTGGTTCAGGGTATNTTGGA 915

Qy   1295 ACGCAAAGTGCCGGAACGCTTTAACGCGCTTCGTGAAAAGCAAATATCGGATTATGAA 1352
          | || |||   || ||       ||      | | | ||| ||||    | |  |  |||
Db    914 ANGCCAAGCAGCGAAAAANGCNAANCTTTTCCTTCAAAGGCAACAGGCCGGGTTGGAA 857


RESULT  10
US-08-425-069-1
; Sequence 1, Application US/08425069
; Patent No. 5728810
;  GENERAL INFORMATION:
;    APPLICANT:  Lewis, Randolph V.
;    APPLICANT:  Xu, Ming
;    APPLICANT:  Hinman, Michael B.
;    TITLE OF INVENTION:  ISOLATED DNA CODING FOR SPIDER SILK
;    TITLE OF INVENTION:  PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
;    TITLE OF INVENTION:  CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF
;    NUMBER OF SEQUENCES:  69
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Birch, Stewart, Kolasch & Birch
;      STREET:  301 No. 5728810th Washington Street
;      CITY:  Falls Church
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22046
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/425,069
;      FILING DATE:  19-APR-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Murphy Jr., Gerald M
;      REGISTRATION NUMBER:  28,977
;      REFERENCE/DOCKET NUMBER:  1447-106P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 205-8000
;      TELEFAX:  (703) 205-8050
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2338 base pairs
;      TYPE:  nucleic acid
```

```
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  Nephilia clavipes
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..2154
;       OTHER INFORMATION:  /product= "Nephila clavipes
;       OTHER INFORMATION:  dragline silk protein"
;     PUBLICATION INFORMATION:
;       AUTHORS:  Xu, Ming
;       AUTHORS:  Lewis, Randolph V.
;       TITLE:  Structure of a protein superfiber:  Spider
;       TITLE:  drafline silk
;       JOURNAL:  Proc. Natl. Acad. Sci. U.S.A.
;       VOLUME:  87
;       PAGES:  7120-7124
;       DATE:  Sept.-1990
;       RELEVANT RESIDUES IN SEQ ID NO:  1: FROM 1 TO 2338
US-08-425-069-1


  Query Match             2.0%;  Score 31.2;  DB 2;  Length 2338;
  Best Local Similarity   46.1%;  Pred. No. 4.8;
  Matches  105;  Conservative    0;  Mismatches  123;  Indels    0;  Gaps    0;

Qy    927 CATGCCCTGCACCAGATGAATAACGTACAACTGGTACATAATGCAGAAAAAGGGGAATAC 986
          || |  | ||| | | || |  |  |  | ||    | || |  || |  | ||  |
Db    320 CAGGTGCAGCAGCCGCTGCAGCAGCTGGAGGTGCCGGACAAGGAGGATATGGAGGCCTTG 379

Qy    987 GACAACAATCTCCAGGGGCTGGTTTCCACGGGGCGTGAGATGTTCCGCCTGGCAACACTG 1046
          || | |||  | | ||    |||   || || ||  ||  |  |  ||||    |||
Db    380 GAAACCAAGGTGCTGGACGAGGTGGACAAGGTGCAGCAGCAGCAGCAGCTGGAGGTGCTG 439

Qy   1047 GAACAGATTGCCCGGGAAAAAGCCGGAACACTGGCTTTAGTCGATGACGTTGAGGTCTAT 1106
          ||  ||   |   |||       | | |   ||     | ||| |  | | ||||  |
Db    440 GACAAGGAGGATATGGAGGTCTTGGAAGCCAAGGTGCAGGACGAGGTGGATTAGGTGGAC 499

Qy   1107 CTGGCGTTCCAGAATAAGCTGAAGGAATCACTTGAGCTGACCAGCGTG 1154
             || |     | |  ||| | || |  |    | ||||  ||  | |
Db    500 AAGGTGCAGGTGCAGCAGCAGCAGCAGCCGGAGGTGCTGGACAAGGCG 547


RESULT  11
US-08-425-069-3
; Sequence 3, Application US/08425069
; Patent No. 5728810
;  GENERAL INFORMATION:
;    APPLICANT:  Lewis, Randolph V.
;    APPLICANT:  Xu, Ming
;    APPLICANT:  Hinman, Michael B.
;    TITLE OF INVENTION:  ISOLATED DNA CODING FOR SPIDER SILK
;    TITLE OF INVENTION:  PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
;    TITLE OF INVENTION:  CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF
;    NUMBER OF SEQUENCES:  69
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Birch, Stewart, Kolasch & Birch
;      STREET:  301 No. 5728810th Washington Street
;      CITY:  Falls Church
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22046
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/425,069
;      FILING DATE:  19-APR-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Murphy Jr., Gerald M
;      REGISTRATION NUMBER:  28,977
;      REFERENCE/DOCKET NUMBER:  1447-106P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 205-8000
;      TELEFAX:  (703) 205-8050
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1995 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    IMMEDIATE SOURCE:
;      CLONE:  p6B
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1785
US-08-425-069-3


  Query Match             2.0%;  Score 31;  DB 2;  Length 1995;
  Best Local Similarity   50.3%;  Pred. No. 5.1;
  Matches   76;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy    913 AGGACCGGGTCACACATGCCCTGCACCAGATGAATAACGTACAACTGGTACATAATGCAG 972
          |||| | || ||   | |||| | |  | ||| |  |||| |   ||| ||  | |
Db    237 AGGATCTGGACAACAAGGCCCAGGAGGATATGGACCACGTCAACAAGGTCCAGGAGGTTA 296

Qy    973 AAAAAGGGGAATACGACAACAATCTCCAGGGGCTGGTTTCCACGGGGCGTGAGATGTTCC 1032
              |  |||  || |  || ||||  |   |   ||      ||  || | ||  |   |
Db    297 TGGACAAGGACAACAAGGACCATCTGGACCAGGCAGTGCAGCCGCAGCCTCAGCCGCAGC 356

Qy   1033 GCCTGGCAACACTGGAACAGATTGCCCGGGA 1063
             | |   |  |||||  | |  |||| |||
Db    357 CTCAGCAGAATCTGGACAACAAGGCCCAGGA 387


RESULT  12
US-08-464-148-3
; Sequence 3, Application US/08464148
; Patent No. 5710026
;  GENERAL INFORMATION:
;    APPLICANT:  Sprecher, Cindy A.
;    TITLE OF INVENTION:  CYTOPLASMIC ANTIPROTEINASE-2 AND
;    TITLE OF INVENTION:  CYTOPLASMIC ANTIPROTEINASE-3 AND CODING SEQUENCES
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend Khourie and Crew
;      STREET:  Steuart Street Tower, One Market Plaza
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  US
;      ZIP:  94105-1493
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/464,148
;      FILING DATE:  05-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/385,500
;      FILING DATE:  08-FEB-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parmelee, Steven W.
;      REGISTRATION NUMBER:  31,990
;      REFERENCE/DOCKET NUMBER:  13952-21
```

```
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 467-9600
;      TELEFAX:  (415) 543-5043
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1393 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  112..1239
;      OTHER INFORMATION:  /product= "CYTOPLASMIC
;      OTHER INFORMATION:  ANTIPROTEINASE-3 PROTEIN"
US-08-464-148-3


  Query Match             2.0%;  Score 30.6;  DB 2;  Length 1393;
  Best Local Similarity   52.8%;  Pred. No. 5.5;
  Matches   66;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy    1009 TTTCCACGGGGCGTGAGATGTTCCGCCTGGCAACACTGGAACAGATTGCCCGGGAAAAAG 1068
           | ||| | |  |  |  ||| | | ||| |   ||  || | | |  ||     | |  |
Db     215 TCTCCTCTGCCCTGGCCATGGTTCTCCTAGGGGCAAAGGGAAACACCGCAACCCAGATGG 274

Qy    1069 CCGGAACACTGGCTTTAGTCGATGACGTTGAGGTCTATCTGGCGTTCCAGAATAAGCTGA 1128
           ||    ||||| ||||| |   || | ||  |  ||| ||| ||||||     || |
Db     275 CCCAGGCACTGTCTTTAAACACAGAGGAAGACATTCATCGGGCTTTCCAGTCGCTTCTCA 334

Qy    1129 AGGAA 1133
             |||
Db     335 CTGAA 339


RESULT  13
US-08-385-500-3
; Sequence 3, Application US/08385500
; Patent No. 5712117
;  GENERAL INFORMATION:
;    APPLICANT:  Sprecher, Cindy A.
;    TITLE OF INVENTION:  CYTOPLASMIC ANTIPROTEINASE-2 AND
;    TITLE OF INVENTION:  CYTOPLASMIC ANTIPROTEINASE-3 AND CODING SEQUENCES
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend Khourie and Crew
;      STREET:  Steuart Street Tower, One Market Plaza
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  US
;      ZIP:  94105-1493
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/385,500
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parmelee, Steven W.
;      REGISTRATION NUMBER:  31,990
;      REFERENCE/DOCKET NUMBER:  13952-21
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 467-9600
;      TELEFAX:  (415) 543-5043
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1393 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  112..1239
;      OTHER INFORMATION:  /product= "CYTOPLASMIC
;      OTHER INFORMATION:  ANTIPROTEINASE-3 PROTEIN"
US-08-385-500-3


  Query Match             2.0%;  Score 30.6;  DB 2;  Length 1393;
  Best Local Similarity   52.8%;  Pred. No. 5.5;
  Matches   66;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy    1009 TTTCCACGGGGCGTGAGATGTTCCGCCTGGCAACACTGGAACAGATTGCCCGGGAAAAAG 1068
           | ||| | |  |  |  ||| | | ||| |   ||  || | | |  ||     | |  |
Db     215 TCTCCTCTGCCCTGGCCATGGTTCTCCTAGGGGCAAAGGGAAACACCGCAACCCAGATGG 274

Qy    1069 CCGGAACACTGGCTTTAGTCGATGACGTTGAGGTCTATCTGGCGTTCCAGAATAAGCTGA 1128
           ||    ||||| ||||| |   || | ||  |  ||| ||| ||||||     || |
Db     275 CCCAGGCACTGTCTTTAAACACAGAGGAAGACATTCATCGGGCTTTCCAGTCGCTTCTCA 334

Qy    1129 AGGAA 1133
             |||
Db     335 CTGAA 339


RESULT  14
US-08-846-784-3
; Sequence 3, Application US/08846784
; Patent No. 5747645
;  GENERAL INFORMATION:
;    APPLICANT:  Sprecher, Cindy A.
;    TITLE OF INVENTION:  CYTOPLASMIC ANTIPROTEINASE-2 AND
;    TITLE OF INVENTION:  CYTOPLASMIC ANTIPROTEINASE-3 AND CODING SEQUENCES
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend Khourie and Crew
;      STREET:  Steuart Street Tower, One Market Plaza
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  US
;      ZIP:  94105-1493
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/846,784
;      FILING DATE:  30-APR-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/385,500
;      FILING DATE:  08-FEB-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parmelee, Steven W.
;      REGISTRATION NUMBER:  31,990
;      REFERENCE/DOCKET NUMBER:  13952-21
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 467-9600
;      TELEFAX:  (415) 543-5043
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1393 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  112..1239
;      OTHER INFORMATION:  /product= "CYTOPLASMIC
;      OTHER INFORMATION:  ANTIPROTEINASE-3 PROTEIN"
```

US-08-846-784-3

Query Match 2.0%; Score 30.6; DB 2; Length 1393;
Best Local Similarity 52.8%; Pred. No. 5.5;
Matches 66; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

```
Qy   1009 TTTCCACGGGGCGTGAGATGTTCCGCCTGGCAACACTGGAACAGATTGCCCGGGAAAAAG 1068
          | |||| | |  |  |  ||| | | ||| |   ||  || | | |  ||     | | |
Db    215 TCTCCTCTGCCCTGGCCATGGTTCTCCTAGGGGCAAAGGGAAACACCGCAACCCAGATGG 274

Qy   1069 CCGGAACACTGGCTTTAGTCGATGACGTTGAGGTCTATCTGGCGTTCCAGAATAAGCTGA 1128
          ||    ||||| |||||  |   || |  ||  |  ||| ||| ||||||     || |
Db    275 CCCAGGCACTGTCTTTAAACACAGAGGAAGACATTCATCGGGCTTTCCAGTCGCTTCTCA 334

Qy   1129 AGGAA 1133
          |||
Db    335 CTGAA 339
```

RESULT 15
PCT-US96-05320A-362/c
; Sequence 362, Application PC/TUS9605320A
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences
; APPLICANT: 9410 Key West Avenue
; APPLICANT: Rockville, MD 20850
; APPLICANT: United States of America
; APPLICANT: Johns Hopkins University
; APPLICANT: 720 Rutland Avenue
; APPLICANT: Baltimore, MD 21205
; APPLICANT: United States of America
; APPLICANT: Mark D. Adams
; APPLICANT: Owen White
; APPLICANT: Hamilton O. Smith
; APPLICANT: J. Craig Venter
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus Influenzae Rd Genome,
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20003-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/05320A
; FILING DATE: April22, 1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/476,102
; FILING DATE: June 7, 1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,429
; FILING DATE: June 7, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Eric K. Steffe
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.014PC01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 362:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 846 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear

PCT-US96-05320A-362

Query Match 2.0%; Score 30.6; DB 5; Length 846;
Best Local Similarity 52.8%; Pred. No. 4;
Matches 66; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

```
Qy    853 AAGATGCTGCGCTGAGAGCAAAAACCTTTGCCATGGCAACAGAGGCAACATCAACCTGCG 912
          | | ||  || |   | | ||  ||| | || ||  | | |   ||  |  ||   |||
Db    634 ATGCTGTCGCACCTGGGGTAACGACCATCGCAATCACTAAAATCGCGCCTACAGTTTGCA 575

Qy    913 AGGACCGGGTCACACATGCCCTGCACCAGATGAATAACGTACAACTGGTACATAATGCAG 972
          |||       |||||| || ||  |  | |||| ||   |   |   | ||||| ||
Db    574 AGGCAGCAACCACACAAGCACTCAATAATGTGAAAAATAAAATTTTATAAAATAATGGAG 515

Qy    973 AAAAA 977
          ||| |
Db    514 AAAGA 510
```

Search completed: October 5, 1999, 18:01:53
Job time: 11553 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 1999, 17:59:29 ; Search time 223.92 Seconds
(without alignments)
1718.450 Million cell updates/sec

Title: US-09-068-804-13
Perfect score: 1538
Sequence: 1 ATGCGTGATTGCCTGAATAA.....CCCGGTGGCGTCTTAACTGA 1538

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		1523.8	99.1	3390	1	T67036	Salmonella secrete
2		142.6	9.3	110000	1	V30458_1	Continuation (2 of
3		142.6	9.3	110000	1	V30459_1	Continuation (2 of
4		39	2.5	280	1	Q61163	Human brain Expres
5	c	37.2	2.4	1150	1	T03650	31-O-desmethyl-FK5
6		36.4	2.4	1355	1	T44494	Murine cyclin-depe
7		35.6	2.3	1758	1	Q42419	IDUA. Isolated nuc
8		35.6	2.3	2155	1	Q42418	IDUA - exons I and
9	c	35	2.3	30001	1	T61016	Total DNA sequence
10	c	35	2.3	30001	1	X05110	S. aureofaciens DN
11	c	33.6	2.2	1868	1	T59712	Endotoxin CryIa(c)
12		33.2	2.2	53789	1	V21187	Amycolatopsis medi
13		32.4	2.1	2478	1	T67287	Soluble starch syn
14		32.4	2.1	2380	1	V66834	Zea mays soluble s
15	c	31.8	2.1	2945	1	Q87790	S. lavendulae mcr
16		31.8	2.1	53577	1	T18551	Human polycystic k
17	c	31.8	2.1	1256	1	T91524	Mycobacterium tube
18	c	31.8	2.1	1256	1	T91458	Mycobacterium tube
19	c	31.8	2.1	16885	1	T33535	BCG deletion regio
20		31.8	2.1	53577	1	T94108	Human PKD1 locus b
21		31.8	2.1	53526	1	T94101	Human PKD1 gene. H
22		31.8	2.1	969	1	V20813	Human haematopoiet
23	c	31.8	2.1	1573	1	V44050	Wheat catalase cDN
24	c	31.8	2.1	10357	1	V52324	Streptococcus pneu
25	c	31.8	2.1	1616	1	V44398	Mycobacterium tube
26	c	31.8	2.1	1616	1	V64507	M. tuberculosis im
27		31.8	2.1	993	1	V63258	Degnerate sequence
28	c	31.8	2.1	1200	1	V63939	Mycobacterium tube
29		31.6	2.1	2934	1	V41922	Nucleotide sequenc
30	c	31.6	2.1	977	1	X16296	Human CD53-like tr
31	c	31.4	2.0	5821	1	T58686	DNA encoding S. fr
32		31.4	2.0	3797	1	T80763	Staphylococcus aur
33		31.4	2.0	3797	1	V53407	DNA encoding a Sta
34		31.4	2.0	25580	1	X12956	Enterococcus faeca
35		31.4	2.0	1002	1	X24498	Human SR-BI gene p
36		31.4	2.0	1002	1	X24562	Human SR-BI gene p
37		31.4	2.0	1002	1	X24590	Human SR-BI gene p
38		31.4	2.0	1002	1	X24654	Human SR-BI gene p
39	c	31.2	2.0	4214	1	N70916	Sequence encoding
40		31.2	2.0	2338	1	V23249	Nephila clavipes s
41	c	31.2	2.0	6761	1	X20517	Polynucleotide seq
42		31.2	2.0	114955	1	X53491	Human adenosine A1
43		31	2.0	1995	1	Q14184	N.clavipes draglin
44		31	2.0	1995	1	V23250	Nephila clavipes s
45		31	2.0	793	1	X39688	Gastric cancer ass

ALIGNMENTS

```
RESULT   1
T67036
ID   T67036 standard; DNA; 3390 BP.
AC   T67036;
DT   04-AUG-1997  (first entry)
DE   Salmonella secreted protein SspH gene.
KW   Salmonella secreted protein; SspH; bacterial-mediated endocytosis;
KW   diagnosis; therapy; vaccine; attenuation; virulence; ss.
OS   Salmonella typhimurium.
FH   Key             Location/Qualifiers
FT   misc_difference 13. .14
FT                   /*tag=  a
FT                   /note= "bases 13-14 are illegible in Fig 29"
FT   misc_difference 37. .38
FT                   /*tag=  b
FT                   /note= "bases 37-38 are illegible in Fig 29"
FT   cds             1698. .3236
FT                   /*tag=  c
FT                   /note= "(Claim 13) translated sequence does not
FT                   fully correspond to SspH sequence of Fig
FT                   27"
PN   WO9718225-A1.
PD   22-MAY-1997.
PF   14-NOV-1996; U18504.
PR   14-NOV-1995; US-006733.
PA   (GEHO ) GEN HOSPITAL CORP.
PI   Miller SI;
DR   WPI; 97-289217/26.
DR   P-PSDB; W15274.
PT   New isolated Salmonella secreted proteins and related genes - used
PT   to develop products for the detection, treatment or prevention of
PT   Salmonella infections
PS   Claim 13; Fig 29; 95pp; English.
CC   A DNA molecule (T67036) includes an open reading frame coding for
CC   Salmonella secreted protein SspH (W15274) plus 5' and 3' flanking
CC   regions.  Ssp polypeptides (see also W15268-71) and nucleic acids
CC   (see also T67028-31) can be used to develop products for the
CC   detection, treatment or prevention of Salmonella infections.
SQ   Sequence   3390 BP;    890 A;    812 C;    939 G;    745 T;
```

Query Match 99.1%; Score 1523.8; DB 1; Length 3390;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1536; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

```
Qy       1 ATGCGTGATTGCCTGAATAACGGCAATCCAGTGCTTAACGTGGGAGCGTCAGGTCTTACC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1698 ATGCGTGATTGCCTGAATAACGGCAATCCAGTGCTTAACGTGGGAGCGTCAGGTCTTACC 1757

Qy      61 ACCTTACCAGACCGTTTACCACCGCATATTACAACACTGGTTATTCCTGATAATAATCTG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1758 ACCTTACCAGACCGTTTACCACCGCATATTACAACACTGGTTATTCCTGATAATAATCTG 1817

Qy     121 ACCAGCCTGCCGGAGTTGCCGGAAGGACTACGGGAGCTGGAGGTCTCTGGTAACCTACAA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1818 ACCAGCCTGCCGGAGTTGCCGGAAGGACTACGGGAGCTGGAGGTCTCTGGTAACCTACAA 1877

Qy     181 CTGACCAGCCTGCCATCGCTGCCGCAGGGACTACAGAAGCTGTGGGCCTATAATAATTGG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1878 CTGACCAGCCTGCCATCGCTGCCGCAGGGACTACAGAAGCTGTGGGCCTATAATAATTGG 1937

Qy     241 CTGGCCAGCCTGCCGACGTTGCCGCCAGGACTAGGGGATCTGGCGGTCTCTAATAACCAG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1938 CTGGCCAGCCTGCCGACGTTGCCGCCAGGACTAGGGGATCTGGCGGTCTCTAATAACCAG 1997

Qy     301 CTGACCAGCCTGCCGGAGATGCCGCCAGCACTACGGGAGCTGAGGGTCTCTGGTAACAAC 360
```

```
Db   1998 CTGACCAGCCTGCCGGAGATGCCGCCAGCACTACGGGAGCTGAGGGTCTCTGGTAACAAC 2057

Qy    361 CTGACCAGCTGCGCGCGCTGCCGTCAGGACTACAGAAGCTGTGGGCCTATAATAATCGGC 420
Db   2058 CTGACCAGCTGCGCGCGCTGCCGTCAGGACTACAGAAGCTGTGGGCCTATAATAATCGGC 2117

Qy    421 TGACCAGCCTGCCGGAGATGTCGCCAGGACTACAGGAGCTGGATGTCTCTCATAACCAGC 480
Db   2118 TGACCAGCCTGCCGGAGATGTCGCCAGGACTACAGGAGCTGGATGTCTCTCATAACCAGC 2177

Qy    481 TGACCCGCCTGCCGCAAAGCCTCACGGGTCTGTCTTCAGCGGCACGCGTATATCTGGACG 540
Db   2178 TGACCCGCCTGCCGCAAAGCCTCACGGGTCTGTCTTCAGCGGCACGCGTATATCTGGACG 2237

Qy    541 GGAATCCACTGTCTGTACGCACTCGTGACAGGCTCTGCGGACATCATTGGCCATTCAGGC 600
Db   2238 GGAATCCACTGTCTGTACGCACTCGTGACAGGCTCTGCGGACATCATTGGCCATTCAGGC 2297

Qy    601 ATCAGGATACACTTCGATATGGCGGGGCCTTCCGTCCCCGGGAAGCCCGGGCACTGCACC 660
Db   2298 ATCAGGATACACTTCCATATGGCGGGGCCTTCCGTCCCCGGGAAGCCCGGGCACTGCACC 2357

Qy    661 TGGCGGTCGCTGACTGGCTGACGTCTGCACGGGAGGGGGAAGCGGCCCAGGCAGACAGAT 720
Db   2358 TGGCGGTCGCTGACTGGCTGACCTCTGCACGGGAGGGGGAAGCGGCCCAGGCAGACAGAT 2417

Qy    721 GGCAGGCGTTCGGACTGGAAGATAACGCCGCCGCCTTCAGCCTGGTCCTGGACAGACTGC 780
Db   2418 GGCAGGCGTTCGGACTGGAAGATAACGCCGCCGCCTTCAGCCTGGTCCTGGACAGACTGC 2477

Qy    781 GTGAGACGGAAAACTTCAAAAAAGACGCGGGCTTTAAGGCACAGATATCATCCTGGCTGA 840
Db   2478 GTGAGACGGAAAACTTCAAAAAAGACGCGGGCTTTAAGGCACAGATATCATCCTGGCTGA 2537

Qy    841 CACAACTGGCTGAAGATGCTGCGCTGAGAGCAAAAACCTTTGCCATGGCAACAGAGGCAA 900
Db   2538 CACAACTGGCTGAAGATGCTGCGCTGAGAGCAAAAACCTTTGCCATGGCAACAGAGGCAA 2597

Qy    901 CATCAACCTGCGAGGACCGGGTCACACATGCCCTGCACCAGATGAATAACGTACAACTGG 960
Db   2598 CATCAACCTGCGAGGACCGGGTCACACATGCCCTGCACCAGATGAATAACGTACAACTGG 2657

Qy    961 TACATAATGCAGAAAAAGGGGAATACGACAACAATCTCCAGGGGCTGGTTTCCACGGGGC 1020
Db   2658 TACATAATGCAGAAAAAGGGGAATACGACAACAATCTCCAGGGGCTGGTTTCCACGGGGC 2717

Qy   1021 GTGAGATGTTCCGCCTGGCAACACTGGAACAGATTG-CCCGGGAAAAAGCCGGAACACTG 1079
Db   2718 GTGAGATGTTCCGCCTGGCAACACTGGAACAGATTGCCCCGGGAAAAAGCCGGAACACTG 2777

Qy   1080 GCTTTAGTCGATGACGTTGAGGTCTATCTGGCGTTCCAGAATAAGCTGAAGGAATCACTT 1139
Db   2778 GCTTTAGTCGATGACGTTGAGGTCTATCTGGCGTTCCAGAATAAGCTGAAGGAATCACTT 2837

Qy   1140 GAGCTGACCAGCGTGACGTCAGAAATGCGTTTCTTTGACGTTTCCGGCGTGACGGTTTCA 1199
Db   2838 GAGCTGACCAGCGTGACGTCAGAAATGCGTTTCTTTGACGTTTCCGGCGTGACGGTTTCA 2897

Qy   1200 GACCTTCAGGCTGCGGACGTTCAGGTGAAAACCGCTGAAAACAGCGGGTTCAGTAAATGG 1259
Db   2898 GACCTTCAGGCTGCGGACGTTCAGGTGAAAACCGCTGAAAACAGCGGGTTCAGTAAATGG 2957

Qy   1260 ATACTGCAGTGGGGGCCGTTACACAGCGTGCTGGAACGCAAAGTGCCGGAACGCTTTAAC 1319
Db   2958 ATACTGCAGTGGGGGCCGTTACACAGCGTGCTGGAACGCAAAGTGCCGGAACGCTTTAAC 3017

Qy   1320 GCGCTTCGTGAAAAGCAAATATCGGATTATGAAGACACGTACCGGAAGCTGTATGACGAA 1379
Db   3018 GCGCTTCGTGAAAAGCAAATATCGGATTATGAAGACACGTACCGGAAGCTGTATGACGAA 3077

Qy   1380 GTGCTGAAATCGTCCGGGCTGGTCGACGATACCGATGCAGAACGTACTATCGGAGTAAGT 1439
Db   3078 GTGCTGAAATCGTCCGGGCTGGTCGACGATACCGATGCAGAACGTACTATCGGAGTAAGT 3137

Qy   1440 GCGATGGATAGTGCGAAAAAAGAATTTCTGGATGGCCTGCGCGCTCTTGTGGATGAGGTG 1499
Db   3138 GCGATGGATAGTGCGAAAAAAGAATTTCTGGATGGCCTGCGCGCTCTTGTGGATGAGGTG 3197

Qy   1500 CTGGGTAGCTATCTGACAGCCCGGTGGCGTCTTAACTGA 1538
Db   3198 CTGGGTAGCTATCTGACAGCCCGGTGGCGTCTTAACTGA 3236
```

RESULT 2
V30458_1
Continuation (2 of 6) of V30458 from base 100001 (Rhizobium species plasmid pNGR234a.
WP Sequence split into 6 fragments LOCUS V30458 Accession V30458

WP	Fragment Name	Begin	End
WP	V30458_0	1	110000
WP	V30458_1	100001	210000
WP	V30458_2	200001	310000
WP	V30458_3	300001	410000
WP	V30458_4	400001	510000
WP	V30458_5	500001	534720

Query Match 9.3%; Score 142.6; DB 1; Length 110000;
Best Local Similarity 50.4%; Pred. No. 9e-32;
Matches 609; Conservative 0; Mismatches 559; Indels 41; Gaps 9;

```
Qy    256 ACGTTGCCGCCAGGACTAGGGGATCTGGCGGTCTCTAATAACCAGCTGACCAGCCTGCCG 315
Db  25173 ACCCTCCCGGGAACGCTTCTGGAGCTGGAGGCCAGCGAGAACCGGCTGACCCGCCTGCCC 25232

Qy    316 GAGATGCCGCCAGCACTACGGGAGCTGAGGGTCTCTGGTAACAACCTGACCAGCTGCGCG 375
Db  25233 GACCTTCCTGCCGGGCTCCAGAGGCTGAACGTCGAAAACAACCGGCTGACCAATCTCCCC 25292

Qy    376 ----CGCTGCCGTCAGGACTACAGAAGCTGTGGGCCTATAATAATCGGCTGACCAGCCTG 431
Db  25293 GAGCCCCTCCCGGCCGCGCTCGAGTGGCTGGGTGCTGGCTACAACCAGCTAACCCGCCTG 25352

Qy    432 CCGGAGATG---TCGCCAGGACTACAGGAGCTGGATGTCTCTCATAACCAGCTGACCCGC 488
Db  25353 CCCGAGATGATTCCCCCCGAGCTTATATGGCTGGGCGCCAGAAACAATCAGCTGACCAGC 25412

Qy    489 CTGCCGCAAAGCCTCACGGGTCTGTCTTCAGCGGCACGCGTATATCTGGACGGGAATCCA 548
Db  25413 GTGCCCGAAAGCCTGCTGACGCAG--CTAGGCCAATGGTCTAGCATTGATTTGGAAAATA 25470

Qy    549 CTGTCTGTACGCACTCGTGACAGGCTCTGCGGACATCATTGGCCATTCAGGCATCAGGAT 608
Db  25471 ATCCGCTGCCGCATGGGGTACAAACAAACTTGGTGACA---GCCATGCATGCCGCGGGTT 25527

Qy    609 ACACTTCGATATGGCGGGGCCTTCCGTCCCCGGGAAGCCCGGGCACTGCAC--CTGGCGG 666
Db  25528 ACGCA---GGACCGCAGATCTTTTTGCCTATGGGACCGGTGGAACTTGCACGACGGCCCC 25584

Qy    667 TCGCTGACTGGCTGACGTCTGCACGGGAGGGGGAAGCGGCCCAGGCAGACAGATGGCAGG 726
Db  25585 TTCACGAGGTTGTCGCGGATTGGCTCGAGGGGGATTTGGAAACGGTGGCCGCCTGGCGGG 25644

Qy    727 CGTTCGGACTGGAAGATAACGCCGCCGCCTTCAGCCTGGTCCTGGACAGACTGCGTGAGA 786
Db  25645 GCTTCGCCAACGAACAAGGTGCCCGGGATTATGCACATTTCCTCGACCGGCTCCGTACCA 25704

Qy    787 CGGAAAACTTCAAAAAAGACGCGGGCTTTAAGGCACAGATATCATCCTGGCTGACACAAC 846
Db  25705 CCGTGAACTATGGCAATGATGCGTTCCGGCAGGC---GGTGGCCATAGGTCTGCGGCAGG 25761

Qy    847 TGGCTGAAGATGCTGCGCTGAGAGCAAAAACCTTTGCCATGGCAACAGAGGCAACATCAA 906
Db  25762 CAGTGGCCAGGCCGCAATTGCGCGCACAGTATTTTGAGCAGGCTTCCGGAGCTAGCGATA 25821

Qy    907 CCTGCGAGGACCGGGTCACACATGCCCTGCACCAGATGAATAACGTACAACTGGTACATA 966
```

```
Db  25822 GCTGTGAGGATCGCATTACTTTGACCTGGAATGGAATGCAGACCGCACTTCTGATCGCTG 25881

Qy    967 ATGCAGAAAAAGGGGAATACGACAACAATCTCCAGGGGCTGGTTTCCACGGGGCGTGAGA 1026
Db  25882 ATGTAGAGGACGGGGTCTATGACGGGTCGCTCCACCAGCTTCTCCAGCACGGCCGTGTCA 25941

Qy   1027 TGTTCCGCCTGGCAACACTGGAACAGATTGCCCGGGAAAAAGCCGGAACACTG------- 1080
Db  25942 TGTTCCGTTTGGAGGCACTGGACGGGATCGCGCGCGAGACGGTCAACTCGCTTCGACGTA 26001

Qy   1080 ----------GCTTTAGTCGATGACGTTGAGGTCTATCTGGCGTTCCAGAATAAGCTGA 1128
Db  26002 CCGATCCGGACGCCGACATCGACGAGATCGAGGTCTATCTTGCCTACCAGACACAGCTGC 26061

Qy   1129 AGGAATCACTTGAGCTGACCAGCGTGACGTCAGAAATGCGTTTCTTTGACGTTTCCGGCG 1188
Db  26062 GCGACACCCTGGAGCTGCGGCACGTCGCCCCAGACATGCGCTTCTTGAACGTCTCTCACG 26121

Qy   1189 TGACGGTTTCAGACCTTCAGGCTGCGGACGTTCAGGTGAAAACCGCTGAAAACAGCGGGT 1248
Db  26122 TTACCGAGGAGGACGTTGCCAGAGCTGCCAGTTCCGTGCGGGAACTTGAGGCGAGGGGAT 26181

Qy   1249 TCAGTAAAT---GGATACTGCAGTGGGGGCCGTTACACAGCGTGCTGGAACGCAAAGTGC 1305
Db  26182 TCGGCGAGTATGTGGCAACGCGCTGGCAACCTTGGGAAAGAGTCATGAGGCGCATCGCCC 26241

Qy   1306 CGGAACGCTTTAACGCGCTTCGTGAAAAGCAAATATCGGATTATGAAGACACGTACCGGA 1365
Db  26242 CCGCAAGCCATGCAGCGATGCAGGAGCAGCTCATCGAAGCGATGGGAGAGGAGTTCAGAA 26301

Qy   1366 AGCTGTATGACGAAGTGCTGAAATCGTCCGGGCTGGTCGACGATACCGATGCAGAACGTA 1425
Db  26302 GCCGCTTGGACGAAAAGCTCGCCGAGCACGGCCTGACGGGTGATGCCGATGCCGAACGGG 26361

Qy   1426 CTATCGGAG 1434
Db  26362 TATTCGGAG 26370


RESULT   3
V30459_1
Continuation (2 of 6) of V30459 from base 100001 (Rhizobium species symbiotic plasmid pN
WP Sequence split into 6 fragments  LOCUS V30459 Accession V30459
WP      Fragment Name      Begin     End
WP      V30459_0               1     110000
WP      V30459_1          100001     210000
WP      V30459_2          200001     310000
WP      V30459_3          300001     410000
WP      V30459_4          400001     510000
WP      V30459_5          500001     536165


  Query Match             9.3%;  Score 142.6;  DB 1;  Length 110000;
  Best Local Similarity   50.4%;  Pred. No. 9e-32;
  Matches  609;  Conservative    0;  Mismatches  559;  Indels   41;  Gaps    9;

Qy     256 ACGTTGCCGCCAGGACTAGGGGATCTGGCGGTCTCTAATAACCAGCTGACCAGCCTGCCG 315
Db   25173 ACCCTCCCGGGAACGCTTCTGGAGCTGGAGGCCAGCGAGAACCGGCTGACCCGCCTGCCC 25232

Qy     316 GAGATGCCGCCAGCACTACGGGAGCTGAGGGTCTCTGGTAACAACCTGACCAGCTGCGCG 375
Db   25233 GACCTTCCTGCCGGGCTCCAGAGGCTGAACGTCGAAAACAACCGGCTGACCAATCTCCCC 25292

Qy     376 ----CGCTGCCGTCAGGACTACAGAAGCTGTGGGCCTATAATAATCGGCTGACCAGCCTG 431
Db   25293 GAGCCCCTCCCGGGCCGCGCTCGAGTGGCTGGGTGCTGGCTACAACCAGCTAACCCGCCTG 25352

Qy     432 CCGGAGATG---TCGCCAGGACTACAGGAGCTGGATGTCTCTCATAACCAGCTGACCCGC 488
Db   25353 CCCGAGATGATTCCCCCCGAGCTTATATGGCTGGGCGCCAGAAACAATCAGCTGACCAGC 25412

Qy     489 CTGCCGCAAAGCCTCACGGGTCTGTCTTCAGCGGCACGCGTATATCTGGACGGGAATCCA 548
Db   25413 GTGCCCGAAAGCCTGCTGACGCAG--CTAGGCCAATGGTCTAGCATTGATTTGGAAAATA 25470

Qy     549 CTGTCTGTACGCACTCGTGACAGGCTCTGCGGACATCATTGGCCATTCAGGCATCAGGAT 608
Db   25471 ATCCGCTGCCGCATGGGGTACAAACAAACTTGGTGACA---GCCATGCATGCCGCGGGTT 25527

Qy     609 ACACTTCGATATGGCGGGGCCTTCCGTCCCCGGGAAGCCCGGGCACTGCAC--CTGGCGG 666
Db   25528 ACGCA---GGACCGCAGATCTTTTTGCCTATGGGACCGGTGGAACTTGCACGACGGCCCC 25584

Qy     667 TCGCTGACTGGCTGACGTCTGCACGGGAGGGGGAAGCGGCCCAGGCAGACAGATGGCAGG 726
Db   25585 TTCACGAGGTTGTCGCGGATTGGCTCGAGGGGGATTTGGAAACGGTGGCCGCCTGGCGGG 25644

Qy     727 CGTTCGGACTGGAAGATAACGCCGCCGCCTTCAGCCTGGTCCTGGACAGACTGCGTGAGA 786
Db   25645 GCTTCGCCAACGAACAAGGTGCCCGGGATTATGCACATTTCCTCGACCGGCTCCGTACCA 25704

Qy     787 CGGAAAACTTCAAAAAAGACGCGGGCTTTAAGGCACAGATATCATCCTGGCTGACACAAC 846
Db   25705 CCGTGAACTATGGCAATGATGCGTTCCGGCAGGC---GGTGGCCATAGGTCTGCGGCAGG 25761

Qy     847 TGGCTGAAGATGCTGCGCTGAGAGCAAAAACCTTTGCCATGGCAACAGAGGCAACATCAA 906
Db   25762 CAGTGGCCAGGCCGCAATTGCGCGCACAGTATTTTGAGCAGGCTTCCGGAGCTAGCGATA 25821

Qy     907 CCTGCGAGGACCGGGTCACACATGCCCTGCACCAGATGAATAACGTACAACTGGTACATA 966
Db   25822 GCTGTGAGGATCGCATTACTTTGACCTGGAATGGAATGCAGACCGCACTTCTGATCGCTG 25881

Qy     967 ATGCAGAAAAAGGGGAATACGACAACAATCTCCAGGGGCTGGTTTCCACGGGGCGTGAGA 1026
Db   25882 ATGTAGAGGACGGGGTCTATGACGGGTCGCTCCACCAGCTTCTCCAGCACGGCCGTGTCA 25941

Qy    1027 TGTTCCGCCTGGCAACACTGGAACAGATTGCCCGGGAAAAAGCCGGAACACTG------- 1080
Db   25942 TGTTCCGTTTGGAGGCACTGGACGGGATCGCGCGCGAGACGGTCAACTCGCTTCGACGTA 26001

Qy    1080 ----------GCTTTAGTCGATGACGTTGAGGTCTATCTGGCGTTCCAGAATAAGCTGA 1128
Db   26002 CCGATCCGGACGCCGACATCGACGAGATCGAGGTCTATCTTGCCTACCAGACACAGCTGC 26061

Qy    1129 AGGAATCACTTGAGCTGACCAGCGTGACGTCAGAAATGCGTTTCTTTGACGTTTCCGGCG 1188
Db   26062 GCGACACCCTGGAGCTGCGGCACGTCGCCCCAGACATGCGCTTCTTGAACGTCTCTCACG 26121

Qy    1189 TGACGGTTTCAGACCTTCAGGCTGCGGACGTTCAGGTGAAAACCGCTGAAAACAGCGGGT 1248
Db   26122 TTACCGAGGAGGACGTTGCCAGAGCTGCCAGTTCCGTGCGGGAACTTGAGGCGAGGGGAT 26181

Qy    1249 TCAGTAAAT---GGATACTGCAGTGGGGGCCGTTACACAGCGTGCTGGAACGCAAAGTGC 1305
Db   26182 TCGGCGAGTATGTGGCAACGCGCTGGCAACCTTGGGAAAGAGTCATGAGGCGCATCGCCC 26241

Qy    1306 CGGAACGCTTTAACGCGCTTCGTGAAAAGCAAATATCGGATTATGAAGACACGTACCGGA 1365
Db   26242 CCGCAAGCCATGCAGCGATGCAGGAGCAGCTCATCGAAGCGATGGGAGAGGAGTTCAGAA 26301

Qy    1366 AGCTGTATGACGAAGTGCTGAAATCGTCCGGGCTGGTCGACGATACCGATGCAGAACGTA 1425
Db   26302 GCCGCTTGGACGAAAAGCTCGCCGAGCACGGCCTGACGGGTGATGCCGATGCCGAACGGG 26361

Qy    1426 CTATCGGAG 1434
Db   26362 TATTCGGAG 26370


RESULT   4
Q61163
ID   Q61163 standard; DNA; 280 BP.
AC   Q61163;
```

```
DT   16-MAR-1994  (first entry)
DE   Human brain Expressed Sequence Tag EST01205.
KW   Gene transcription product; genetic markers; tagging; in vivo;
KW   transcription; mapping; locations; chromosomes; chromosomal; ss.
OS   Homo sapiens.
PN   WO9316178-A.
PD   19-AUG-1993.
PF   12-FEB-1993; U01294.
PR   12-FEB-1992; US-837195.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
PI   Adams MD, Moreno RF, Venter CJ;
DR   WPI; 93-272882/34.
PT   Enriched oligonucleotides and corresp. sequences - used as
PT   markers for human genes transcribed in-vivo, facilitate tagging
PT   of most human genes
PS   Example 4; Page 448; 500pp; English.
CC   The Expressed Sequence Tag was isolated from a human brain cDNA
CC   library as part of a large set of ESTs which can be used as markers
CC   for human genes transcribed in vivo. They can be used to facilitate
CC   tagging of most human genes, for mapping locations of expressed genes
CC   on chromosomes, for individual or forensic identification, for mapping
CC   locations of disease-associated genes, for identification of tissue
CC   type, and for prepn. of antisense sequences, probes and constructs.
CC   EST01205 has an "Excellent" coding probability as evaluated using the
CC   coding-region prediction program CRM. See also Q59041-Q61440.
SQ   Sequence   280 BP;    63 A;    67 C;    97 G;    53 T;


  Query Match              2.5%;  Score 39;  DB 1;  Length 280;
  Best Local Similarity   58.0%;  Pred. No. 0.032;
  Matches   69;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy   1047 GAACAGATTGCCCGGGAAAAAGCCGGAACACTGGCTTTAGTCGATGACGTTGAGGTCTAT 1106
            || ||| |   || ||||  |  |   |   ||| || | | || ||||  |||| ||||   |
Db     28 GAGCAGCTCAGCCAGGAAGAGGTTGATGCACAGGTTCTTGTGGATGGTGTTGCGGTCGGT 87

Qy   1107 CTGGCGTTCCAGAATAAGCTGAAGGAATCACTTGAGCTGACCAGCGTGACGTCAGAAAT 1165
            |||  |  || | | ||||  |||||       |  || || |||    |||    || ||
Db     88 CTGCAGCCCCCGCAGAAGCAGAAGGTGGAGATGCAGATGGCCAAGCAGACCAGGGAGAT 146


RESULT   5
T03650/c
ID   T03650 standard; DNA; 1150 BP.
AC   T03650;
DT   16-AUG-1996  (first entry)
DE   31-O-desmethyl-FK506 O:methyltransferase gene FKMT2.
KW   Streptomyces; 31-O-desmethyl-FK506 O:methyltransferase; terminator;
KW   expression cassette; strong promoter; PermE*; ribosome binding site;
KW   translation start site; transcription terminator; selection marker;
KW   hygromycin; integration element; minicircle; chromosome; E.coli;
KW   replicon; ColEI; overexpression; immunosuppressant drug; ss.
OS   Streptomyces sp. strain MA6548.
FH   Key             Location/Qualifiers
FT   cds             62. .844
FT                   /*tag=  a
FT                   /product= 31-O-desmethyl-FK506 O:methyltransferase
FT   repeat_unit     842. .893
FT                   /*tag=  b
FT                   /note= "inverted repeat unit"
FT   repeat_unit     845. .877
FT                   /*tag=  c
FT                   /note= "inverted repeat unit"
PN   WO9600282-A1.
PD   04-JAN-1996.
PF   19-JUN-1995; U07784.
PR   23-JUN-1994; US-264861.
PA   (MERI ) MERCK & CO INC.
PI   Motamedi H,  Shafiee A;
DR   WPI; 96-068866/07.
PT   Integrative and replicative expression vectors for Streptomyces -
PT   esp. for expressing 31-O-desmethyl-FK506 O:methyltransferase for
PT   prodn of the immunosuppressive drug FK506
PS   Claim 23; Page 36; 48pp; English.
CC   This is the nucleotide sequence of the Streptomyces gene FKMT2 encoding
CC   31-O-desmethyl-FK506 O:methyltransferase.  The sequence contains the
CC   coding region for the protein and includes the termination sequence.
CC   The gene was used in the construction of a Streptomyces expression
CC   cassette comprising; (a) a strong promoter e.g. the PermE* constitutive
CC   promoter, (b) a synthetic ribosome binding site, (c) a translation start
CC   site (contained within the restriction enzyme site for NdeI), (d) a
CC   multiple cloning site, (e) a transcription terminator downstream from the
CC   MCS, esp. the terminator region of the FKMT2 gene.  The vector also
CC   contains a selection marker, pref. the hygromycin resistance gene, an
CC   integration element, pref. from a minicircle, for integration into the
CC   Streptomyces chromosome and an E.coli replicon, pref. the ColEI replicon.
CC   The vectors can be used for the integration of single copies of desired
CC   genes into the Streptomyces genome or for the transfer of multiple copies
CC   of genes of interest for overexpression purposes, e.g. for the conversion
CC   of 31-desmethyl-FK506 to the immunosuppressant drug FK506 by integration
CC   of the FKMT2 gene.
SQ   Sequence   1150 BP;    164 A;    418 C;    396 G;    172 T;


  Query Match              2.4%;  Score 37.2;  DB 1;  Length 1150;
  Best Local Similarity   52.6%;  Pred. No. 0.22;
  Matches   81;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy    622 GCGGGGCCTTCCGTCCCCGGGAAGCCCGGGCACTGCACCTGGCGGTCGCTGACTGGCTGA 681
          |||| ||| || | |  |||   | || || ||   ||   | ||  || ||  |  | |
Db   1130 GCGGAGCCCTCGGACATCGGCGGGGCCTGGGACGCGACGCTGGGGCAGCGGATCGCGTCA 1071

Qy    682 CGTCTGCACGGGAGGGGGAAGCGGCCCAGGCAGACAGATGGCAGGCGTTCGGACTGGAAG 741
          |    | |   | | ||  |  | ||  || | | |  | | ||||| || || || ||
Db   1070 CTGACGGAGTCGGAGCGGCGGGTGGCCGCGCTGGCCGCGGTCGGGCGTACGAACAGGGAG 1011

Qy    742 ATAACGCCGCCGCCTTCAGCCTGGTCCTGGACAG 775
          ||  |   || ||  |  | |  | || | || |
Db   1010 ATCGCCGAGCAGCTGTTCGTCACGGCCAGCACGG 977


RESULT   6
T44494
ID   T44494 standard; cDNA; 1355 BP.
AC   T44494;
DT   22-FEB-1997  (first entry)
DE   Murine cyclin-dependent kinase inhibitor p57KIP2 cDNA.
KW   Cyclin-dependent kinase inhibitor; CDK; CDI; p57KIP2; cell cycle;
KW   cancer; hyperplasia; ulcer; hyperproliferation; hypoproliferation;
KW   diagnosis; gene therapy; ss.
OS   Mus sp.
FH   Key             Location/Qualifiers
FT   cds             41. .1087
FT                   /*tag=  a
PN   WO9631534-A1.
PD   10-OCT-1996.
PF   03-APR-1996; U04563.
PR   03-APR-1995; US-415655.
PA   (SLOK ) SLOAN KETTERING INST CANCER RES.
PI   Lee M,  Massague J;
DR   WPI; 96-464971/46.
DR   P-PSDB; W01437.
PT   Mammalian p57-KIP2 and related DNA - used to develop prods. useful
PT   for diagnosis and treatment of hyper- and hypo-proliferative
PT   disorders
PS   Disclosure; Page 64-66; 98pp; English.
CC   A cDNA clone (T44494) codes for murine p57KIP2 (W01437), a novel
CC   member of the p21CIP1/p27KIP1 cyclin-dependent kinase (CDK)
CC   inhibitor family that is a potent inhibitor of G1 and S phase CDKs
CC   and which plays a role in cell cycle control.  A mouse embryo cDNA
CC   library was screened with a mouse p21 cDNA probe.  Restriction
CC   analysis of positive clones provided the novel p57KIP2 species.
CC   p57KIP2 nucleic acids can be used to produce recombinant p57KIP2
CC   in host cells, to design probes useful in detecting mutations of
CC   the p57KIP2 gene, and in the treatment of hyperproliferative
CC   disorders associated with such mutations.  The isolated cDNA was
```

```
CC   used to identify human p57-KIP2 (see also T44499).
SQ   Sequence   1355 BP;    267 A;    405 C;    469 G;    214 T;


  Query Match             2.4%;  Score 36.4;  DB 1;  Length 1355;
  Best Local Similarity   53.5%;  Pred. No. 0.41;
  Matches   76;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy    632 CCGTCCCCGGGAAGCCCGGGCACTGCACCTGGCGGTCGCTGACTGGCTGACGTCTGCACG 691
          ||  |  | ||   ||||| |||  |   |||| || |||| |   | |   ||||  |
Db    349 CCTGCAGCTGGGGCCCCGGCCACCCCCGGTGGCCGTGGCTGTCATCCCGCGTTCTGGGCC 408

Qy    692 GGAGGGGGAAGCGGCCCAGGCAGACAGATGGCAGGCGTTCGGACTGGAAGATAACGCCGC 751
          |  ||  |  | |||||  |  | |  |  | |||||   |  | |      | ||||  |
Db    409 GCCGGCTGGCGAGGCCCCCGACGGCCTAGAGGAGGCGCCTGAGCAGCCGCCCAGCGCCCC 468

Qy    752 CGCCTTCAGCCTGGTCCTGGAC 773
            ||||    | |||||  ||||
Db    469 AGCCTCGGCCGTGGTCGCGGAC 490


RESULT   7
Q42419
ID   Q42419 standard; cDNA; 1758 BP.
AC   Q42419;
DT   27-SEP-1993  (first entry)
DE   IDUA.
KW   Alpha-L-iduronidase; glycosaminoglycan alpha-L-iduronohydrolase;
KW   mucopolysaccharidosis type I; MPS-1; polymorphisms;
KW   Huntington's disease; primer; probe; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   exon            583. .740
FT                   /*tag= a
FT                   /number= I
FT   exon            1307. .1447
FT                   /*tag= b
FT                   /number= II
FT   repeat_unit     143. .298
FT                   /*tag= c
FT                   /note= "Alu repeat sequence"
FT   promoter        435. .448
FT                   /*tag= d
FT                   /note= "potential GC box"
FT   promoter        440. .453
FT                   /*tag= e
FT                   /note= "potential GC box"
FT   promoter        474. .487
FT                   /*tag= f
FT                   /note= "potential GC box"
FT   promoter        525. .537
FT                   /*tag= g
FT                   /note= "potential GC box"
PN   WO9310244-A.
PD   27-MAY-1993.
PF   12-NOV-1992; AU0611.
PR   14-NOV-1991; AU-009490.
PA   (WOME-) WOMEN'S & CHILDREN'S HOSPITAL.
PI   Anson DS,  Clements PR,  Hopwood JJ,  Morris CP,  Nelson PV;
PI   Orsborn AM,  Scott HS;
DR   WPI; 93-182558/22.
PT   Isolated nucleic acid molecule for alpha-L-iduronidase
PT   deficiency diagnosis - comprises sequence of nucleotide(s)
PT   encoding or complementary to mammalian alpha-L-iduronidase
PT   fragment, for treating lysosomal storage disorder
PS   Disclosure; Fig 4A; 47pp; English.
CC   The IDUA genomic sequence is valuable for defining mutations in
CC   mucopolysaccharidosis type I (MPS-1) patients, for defining
CC   diagnostically useful polymorphisms for MPS-1 and Huntington's disease
CC   and for refining the genetic and physical map of the IDUA gene.
CC   The genomic sequence is given in Q42419 and Q45995.  The gene is
CC   split into 14 exons spanning ca. 19 kb.
SQ   Sequence   1758 BP;    270 A;    616 C;    559 G;    313 T;


  Query Match             2.3%;  Score 35.6;  DB 1;  Length 1758;
  Best Local Similarity   60.2%;  Pred. No. 0.8;
  Matches   59;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy    615 CGATATGGCGGGGCCTTCCGTCCCCGGGAAGCCCGGGCACTGCACCTGGCGGTCGCTGAC 674
          |||    ||| |||| | |||||||  |    |||| ||   ||  ||||||  ||   | |
Db    568 CGAGCACGCGTGGCCATGCGTCCCCTGCGCCCCCGCGCCGCGCTGCTGGCGCTCCTGGCC 627

Qy    675 TGGCTGACGTCTGCACGGGAGGGGGAAGCGGCCCAGGC 712
          | |||   | | || |    || ||   |||| ||||
Db    628 TCGCTCCTGGCCGCGCCCCCGGTGGCCCCGGCCGAGGC 665


RESULT   8
Q42418
ID   Q42418 standard; cDNA; 2155 BP.
AC   Q42418;
DT   27-SEP-1993  (first entry)
DE   IDUA - exons I and II.
KW   Alpha-L-iduronidase; glycosaminoglycan alpha-L-iduronohydrolase;
KW   mucopolysaccharidosis type I; MPS-1; primer; probe; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             89. .2050
FT                   /*tag= a
FT   signal_peptide  89. .167
FT                   /*tag= b
FT   mat_peptide     168. .2047
FT                   /*tag= c
FT   exon            90. .246
FT                   /*tag= d
FT                   /number= I
FT   exon            247. .387
FT                   /*tag= e
FT                   /number= II
FT                   /note= "alternatively spliced in some RNA transcripts"
FT   exon            388. .473
FT                   /*tag= f
FT                   /number= III
FT   exon            474. .582
FT                   /*tag= g
FT                   /number= IV
FT                   /note= "alternatively spliced in some RNA transcripts"
FT   exon            583. .2047
FT                   /*tag= h
FT                   /number= V
FT   primer_bind     1. .25
FT                   /*tag= i
FT                   /label= ID58
FT                   /note= "sense oligonucleotide"
FT   primer_bind     97. .100
FT                   /*tag= j
FT                   /label= ID60
FT                   /note= "sense oligonucleotide"
FT   primer_bind     185. .204
FT                   /*tag= k
FT                   /label= ID13
FT                   /note= "oligonucleotide"
FT   primer_bind     223. .246
FT                   /*tag= l
FT                   /label= ID56
FT                   /note= "oligonucleotide"
FT   primer_bind     409. .428
FT                   /*tag= m
FT                   /label= IDNT
FT                   /note= "sense oligonucleotide"
FT   primer_bind     425. .450
FT                   /*tag= n
FT                   /label= ID57
FT                   /note= "antisense oligonucleotide"
FT   primer_bind     612. .631
```

```
FT                   /*tag=  o
FT                   /label= ID39
FT                   /note= "antisense oligonucleotide"
FT   primer_bind     883. .905
FT                   /*tag=  p
FT                   /label= ID61
FT                   /note= "antisense oligonucleotide"
PN   WO9310244-A.
PD   27-MAY-1993.
PF   12-NOV-1992; AU0611.
PR   14-NOV-1991; AU-009490.
PA   (WOME-) WOMEN'S & CHILDREN'S HOSPITAL.
PI   Anson DS,  Clements PR,  Hopwood JJ,  Morris CP,  Nelson PV;
PI   Orsborn AM,  Scott HS;
DR   WPI; 93-182558/22.
DR   P-PSDB; R37425.
PT   Isolated nucleic acid molecule for alpha-L-iduronidase
PT   deficiency diagnosis - comprises sequence of nucleotide(s)
PT   encoding or complementary to mammalian alpha-L-iduronidase
PT   fragment, for treating lysosomal storage disorder
PS   Claim 9; Fig 2; 47pp; English.
CC   The IDUA nucleic acid was isolated using oligonucleotides
CC   synthesised on the basis of IDUA amino acid sequences.  The nucleic
CC   acid can be used for the treatment of IDUA deficiency, e.g.
CC   mucopolysaccharidosis type I.  It can also be used for screening for
CC   aberrations in the IDUA gene.
SQ   Sequence   2155 BP;    333 A;    864 C;    613 G;    345 T;

  Query Match             2.3%;  Score 35.6;  DB 1;  Length 2155;
  Best Local Similarity   60.2%;  Pred. No. 0.89;
  Matches   59;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy    615 CGATATGGCGGGGCCTTCCGTCCCCGGGAAGCCCGGGCACTGCACCTGGCGGTCGCTGAC 674
          |||    ||| |||| | |||||||  |    |||| ||   ||  |||||| ||   | |
Db     74 CGAGCACGCGTGGCCATGCGTCCCCTGCGCCCCCGCGCCGCGCTGCTGGCGCTCCTGGCC 133

Qy    675 TGGCTGACGTCTGCACGGGAGGGGGAAGCGGCCCAGGC 712
          | |||   | | || |    || ||   ||||| ||||
Db    134 TCGCTCCTGGCCGCGCCCCCGGTGGCCCCGGCCGAGGC 171


RESULT   9
T61016/c
ID   T61016 standard; DNA; 30001 BP.
AC   T61016;
DT   21-APR-1997  (first entry)
DE   Total DNA sequence from cosmid clones LP(2)127 and LP(2)128.
KW   Cosmid clone; LP(2)127; LP(2)128; chlortetracycline;
KW   biosynthetic pathway; recombinant; production; antibiotic;
KW   heterologous host; Streptomyces lividans; ss.
OS   Streptomyces aureofaciens.
PN   US5589385-A.
PD   31-DEC-1996.
PF   26-JUL-1990; 558039.
PR   26-JUL-1990; US-558039.
PR   26-JUL-1990; US-558040.
PR   15-JAN-1992; US-821419.
PR   15-JAN-1992; US-821109.
PR   22-SEP-1993; US-125468.
PA   (AMCY ) AMERICAN CYANAMID CO.
PI   Fantini SE,  Lotvin JA,  Ryan MJ,  Strathy N;
DR   WPI; 97-076853/07.
PT   DNA encoding tetracycline biosynthetic pathway proteins -
PT   specifically from Streptomyces aureofaciens for expression in
PT   heterologous hosts, specifically S. lividans
PS   Claim 3; Columns 17-44; 39pp; English.
CC   The present sequence is the total DNA sequence from cosmid clones
CC   LP(2)127 and LP(2)128, which contains a Streptomyces aureofaciens
CC   (ATCC 13899) DNA that encodes the proteins of the entire
CC   chlortetracycline biosynthetic pathway. The biosynthetic gene,
CC   which can be expressed in heterologous hosts, especially
CC   S. lividans, may be useful in the production of antibiotics.
SQ   Sequence   30001 BP;    4055 A;    10939 C;    11019 G;    3988 T;

  Query Match             2.3%;  Score 35;  DB 1;  Length 30001;
  Best Local Similarity   52.4%;  Pred. No. 4.9;
  Matches   77;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy    621 GGCGGGGCCTTCCGTCCCCGGGAAGCCCGGGCACTGCACCTGGCGGTCGCTGACTGGCTG 680
          | | ||||   ||| |    ||  ||   |   | || ||  |||||| | ||| |||
Db  16130 GCCCGGGCGCACCGCCTTGTGGGCGCGGCGCAGCAGCTCCCTGCGGTCCCCGACCGGCCC 16071

Qy    681 ACGTCTGCACGGGAGGGGGAAGCGGCCCAGGCAGACAGATGGCAGGCGTTCGGACTGGAA 740
          |   | ||| | | |   ||||    ||| |    |    ||||||     ||    |||
Db  16070 AGTCCGGCAGGCGTGACCGAAGATCACCACGTCCGCCTCCGGCAGGTCGGTGGCGAAGAA 16011

Qy    741 GATAACGCCGCCGCCTTCAGCCTGGTC 767
          |  | |||||  |   ||  || ||||
Db  16010 GTCACCGCCGTGGAAGTCCACCCGGTC 15984


RESULT  10
X05110/c
ID   X05110 standard; DNA; 30001 BP.
AC   X05110;
DT   07-APR-1999  (first entry)
DE   S. aureofaciens DNA in cosmid clones LP2-127 and LP2-128.
KW   Chlortetracycline; tetracycline; biosynthetic; actinomycete;
KW   S. ariseofuscus; S. ambofaciens; antibiotic resistance gene;
KW   S lividans; S aureofaciens; cosmid clone; antibiotic; ss.
OS   Streptomyces aureofaciens.
PN   US5866410-A.
PD   02-FEB-1999.
PF   07-JUN-1995; 474933.
PR   22-SEP-1993; US-125468.
PR   26-JUL-1990; US-558039.
PR   26-JUL-1990; US-558040.
PR   15-JAN-1992; US-821109.
PR   15-JAN-1992; US-821419.
PR   07-JUN-1995; US-474933.
PA   (AMCY ) AMERICAN CYANAMID CO.
PI   Fantini SE, Lotvin JA, Ryan MJ, Strathy N;
DR   WPI; 99-141936/12.
PT   Two plasmids for cloning the biosynthetic pathways of tetracycline,
PT   chlortetracycline, and their derivatives - comprise an origin of
PT   replication, cohesive end sites and optionally, an antibiotic
PT   resistance gene
PS   Disclosure; Fig 4A-L; 39pp; English.
CC   The invention relates to combination of two plasmids for cloning the
CC   genes encoding the biosynthetic production pathways of chlortetracycline,
CC   tetracycline, or their derivatives. One plasmid contains an origin of
CC   replication, an actinomycete (such as Streptomyces lividans,
CC   S. ariseofuscus, or S. ambofaciens) active antibiotic resistance gene and
CC   three or more tandem cohesive end sites. The second plasmid contains an
CC   actinomycete active origin of replication and three or more tandem
CC   cohesive end sites. The invention can be used to clone a large amount of
CC   genetic material for the heterologous production of the antibiotics,
CC   chlortetracycline, tetracycline and their analogues. A cluster of genes
CC   can be inserted into the plasmid allowing a biosynthetic pathway to be
CC   transferred in its entirety to a heterologous host. The invention
CC   specifically relates to cloning of the entire tetracycline and
CC   chlortetracycline from Streptomyces aureofaciens and its expression in a
CC   heterologous host such as S. lividans. The present sequence represents a
CC   S. aureofaciens DNA sequence from the cosmid clones designated LP2-127
CC   and LP2-128.
SQ   Sequence   30001 BP;    4055 A;    10939 C;    11019 G;    3988 T;

  Query Match             2.3%;  Score 35;  DB 1;  Length 30001;
  Best Local Similarity   52.4%;  Pred. No. 4.9;
  Matches   77;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy    621 GGCGGGGCCTTCCGTCCCCGGGAAGCCCGGGCACTGCACCTGGCGGTCGCTGACTGGCTG 680
          | | ||||   ||| |    ||  ||   |   | || ||  |||||| | ||| |||
```

```
Db  16130 GCCCGGGCGCACCGCCTTGTGGGCGCGGCGCAGCAGCTCCCTGCGGTCCCCGACCGGCCC 16071

Qy    681 ACGTCTGCACGGGAGGGGGAAGCGGCCCAGGCAGACAGATGGCAGGCGTTCGGACTGGAA 740
Db  16070 AGTCCGGCAGGCGTGACCGAAGATCACCACGTCCGCCTCCGGCAGGTCGGTGGCGAAGAA 16011

Qy    741 GATAACGCCGCCGCCTTCAGCCTGGTC 767
Db  16010 GTCACCGCCGTGGAAGTCCACCCGGTC 15984


RESULT 11
T59712/c
ID   T59712 standard; DNA; 1868 BP.
AC   T59712;
DT   12-MAY-1997 (first entry)
DE   Endotoxin CryIa(c) gene optimised for plant codon usage.
KW   Gene expression cassette; promoter; alcR regulator; insecticide;
KW   CryIA(c); crystal protein; delta-endotoxin; Bacillus thuringinesis;
KW   Lepidoptera; insect resistance; transgenic plant;
KW   crop protection; biological control; alcR; gene switch; ds.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   cds             13. .1163
FT                   /*tag= a
PN   WO9706268-A2.
PD   20-FEB-1997.
PF   29-JUL-1996; G01846.
PR   08-AUG-1995; GB-016241.
PA   (ZENE ) ZENECA LTD.
PI   Jepson I, Paine JAM;
DR   WPI; 97-154272/14.
PT   Chemically inducible expression cassette - contains inducible
PT   promoter activated by alcR regulator in presence of alcohol or
PT   ketone inducer, used for insecticide production in plants
PS   Example 3; Fig 7; 52pp; English.
CC   A CryIA(c) gene (T59712) codes for a Bacillus thuringiensis
CC   endotoxin that is entomocidal to Lepidopteran larvae. It is a
CC   modified synthetic sequence, optimised for plant codon usage. The
CC   gene can be incorporated into novel chemically-inducible plant gene
CC   expression cassettes (see also T59712) in which the CryIA(c) gene
CC   is expressed from an inducible promoter that is activated by an
CC   alcR regulator in the presence of an alcohol or ketone inducer.
CC   Transgenic plants carrying the expression cassette are inducibly
CC   resistant to Lepidopteran pests. Inducible expression reduces the
CC   risk of developing resistant insect strains.
SQ   Sequence  1868 BP;  361 A;  702 C;  491 G;  314 T;


  Query Match             2.2%;  Score 33.6;  DB 1;  Length 1868;
  Best Local Similarity   46.6%;  Pred. No. 3.2;
  Matches  108;  Conservative    0;  Mismatches  124;  Indels    0;  Gaps    0;

Qy    118 CTGACCAGCCTGCCGGAGTTGCCGGAAGGACTACGGGAGCTGGAGGTCTCTGGTAACCTA 177
Db   1097 CCGATCAGCCTGGTCAGGTCGTTGTAGCGGCTGTTGATGGTGGCGGCGTCGAAGCCCCAG 1038

Qy    178 CAACTGACCAGCCTGCCATCGCTGCCGCAGGGACTACAGAAGCTGTGGGCCTATAATAAT 237
Db   1037 CGCTGGCCGAACACGGACTCGATGTAGCCGCGGTTCTGGATGTTGTTGCCGGAGCTGTTC 978

Qy    238 TGGCTGGCCAGCCTGCCGACGTTGCCGCCAGGACTAGGGGATCTGGCGGTCTCTAATAAC 297
Db    977 AGCCTGACGAGGTCGCCGCCGGTGAAGCCCGGGCCGGAGATCACGGAGCCGTTGAACAGG 918

Qy    298 CAGCTGACCAGCCTGCCGGAGATGCCGCCAGCACTACGGGAGCTGAGGGTCT 349
Db    917 AAGTTGCCCTTCACAGCCGGGATCTGGGTGATGCTGTCGGAGGCGATGATGT 866


RESULT 12
V21187
ID   V21187 standard; DNA; 53789 BP.
AC   V21187;
DT   24-JUL-1998 (first entry)
DE   Amycolatopsis mediterranei rifamycin synthesis gene cluster fragment.
KW   Amycolatopsis mediterranei; rifamycin; synthesis; gene cluster;
KW   polyketide synthase; actinomycete; ansamycin; ds.
OS   Amycolatopsis mediterranei.
FH   Key             Location/Qualifiers
FT   CDS             1825. .15543
FT                   /*tag= a
FT                   /label= ORF_A
FT                   /product= "polyketide synthase"
FT   CDS             15550. .30759
FT                   /*tag= b
FT                   /label= ORF_B
FT                   /product= "polyketide synthase"
FT   CDS             30895. .36060
FT                   /*tag= c
FT                   /label= ORF_C
FT                   /product= "polyketide synthase"
FT   CDS             36259. .41325
FT                   /*tag= d
FT                   /label= ORF_D
FT                   /product= "polyketide synthase"
FT   CDS             41373. .51614
FT                   /*tag= e
FT                   /label= ORF_E
FT                   /product= "polyketide synthase"
FT   CDS             51713. .5293
FT                   /*tag= f
FT                   /label= ORF_F
FT                   /product= "polyketide synthase"
PN   WO9807868-A1.
PD   26-FEB-1998.
PF   18-AUG-1997; E04495.
PR   20-AUG-1996; EP-810551.
PA   (NOVS ) NOVARTIS AG.
PI   Engel N, Schupp T, Toupet C;
DR   WPI; 98-169172/15.
DR   P-PSDB; W52845-W52850.
PT   Amycolatopsis mediterranei rifamycin synthesis gene cluster - used
PT   to produce rifamycin and rifamycin analogues
PS   Claim 4; Page 53-102; 205pp; English.
CC   The present sequence represents a Amycolatopsis mediterranei rifamycin
CC   synthesis gene cluster DNA fragment from the present invention. The
CC   DNA fragment comprises a DNA region involved directly or indirectly
CC   in the gene cluster responsible for rifamycin synthesis, including
CC   the adjacent DNA regions to the right and left which, by reason of
CC   their function in connection with rifamycin biosynthesis, qualify
CC   as constituents of this rifamycin gene cluster, and functional
CC   fragments, derivatives or constituents of these. The Amycolatopsis
CC   mediterranei rifamycin synthesis gene cluster DNA fragment can be used
CC   for producing rifamycin, rifamycin analogues or precursors. It can also
CC   be used for inactivating or modifying genes involved in ansamycin or
CC   rifamycin biosynthesis. The DNA can be used for constructing mutant
CC   actinomycetes strains from which the natural rifamycin or ansamycin
CC   biosynthesis gene cluster has been partly or completely deleted. The
CC   DNA fragment can be used for assembling a library of polyketide
CC   synthases, which can be used for assembling a library of polyketides.
CC   A hybridisation probe of the invention can be used for identifying DNA
CC   fragments involved in the biosynthesis of ansamycins.
SQ   Sequence  53789 BP;  6707 A;  19183 C;  20504 G;  7395 T;


  Query Match             2.2%;  Score 33.2;  DB 1;  Length 53789;
  Best Local Similarity   49.4%;  Pred. No. 22;
  Matches   86;  Conservative    0;  Mismatches   88;  Indels    0;  Gaps    0;

Qy    621 GGCGGGGCCTTCCGTCCCCGGGAAGCCCGGGCACTGCACCTGGCGGTCGCTGACTGGCTG 680
Db  38194 GGCCGCGGCGGCATGGCCTCGGTCGCGCTGAGCGAGGAAGACGCGGTCGCGCGGTTGCGG 38253

Qy    681 ACGTCTGCACGGGAGGGGGAAGCGGCCCAGGCAGACAGATGGCAGGCGTTCGGACTGGAA 740
```

```
Db  38254 CACTGGGCGGACCGGGTCGAGGTGGCCGCGGTCAACAGCCCGTCGTCGGTGGTGATCGCC 38313

Qy    741 GATAACGCCGCCGCCTTCAGCCTGGTCCTGGACAGACTGCGTGAGACGGAAAAC 794
          |   ||||||  ||| ||  || || ||| ||   ||||   |    ||| | |
Db  38314 GGCGACGCCGAAGCCCTCGACCAGGCCCTCGAAGCACTGACCGGCCAGGACATC 38367
```

RESULT 13
T67287
ID T67287 standard; cDNA; 2478 BP.
AC T67287;
DT 11-SEP-1997 (first entry)
DE Soluble starch synthase cDNA clone SSS56.
KW Soluble starch synthase; SSS56; transgenic plant; cereal; maize;
KW ss.
OS Zea mays inbred line W64A.
PN WO9720936-A1.
PD 12-JUN-1997.
PF 04-DEC-1996; G02990.
PR 06-DEC-1995; GB-024938.
PA (ZENE) ZENECA LTD.
PI Keeling PL, Knight ME;
DR WPI; 97-319782/29.
PT cDNA encoding soluble starch synthase - used to produce transgenic
PT plants with increased capacity for producing and storing starch
PS Claim 5; Page 23-26; 44pp; English.
CC cDNA clones SSS10.52, SSS6.31 and SSS56 (T67285-87) code for maize
CC soluble starch synthase (SSS). They were isolated from a maize
CC inbred line W64A library by screening with a probe (T67288) based
CC on pea SSS, rice SSS and maize GBSS. The cDNAs can be used to
CC produce plants with an increased capacity for producing starch, or
CC a capacity to produce starch with an altered fine structure. They
CC can also be used to isolate the corresponding genomic sequences
CC from crop plants, to determine the contribution of the SSS gene to
CC the net regulation of starch biosynthesis, and to modify the levels
CC of starch produced by the plant. Transgenic plants (esp. maize)
CC can be used to produce hybid plants which have higher rates of
CC starch synthesis at temperatures above the normal optimum.
SQ Sequence 2478 BP; 347 A; 276 C; 533 G; 289 T;

```
  Query Match             2.1%;  Score 32.4;  DB 1;  Length 2478;
  Best Local Similarity   28.5%;  Pred. No. 8.5;
  Matches   59;  Conservative   31;  Mismatches  117;  Indels    0;  Gaps    0;

Qy    454 AGGAGCTGGATGTCTCTCATAACCAGCTGACCCGCCTGCCGCAAAGCCTCACGGGTCTGT 513
          | |   | |  |  :  |  ||    :|  | :| :     || :|  | | ||   |
Db    737 AYGTNGTNGGNGCNYTNCCNAARGCNYTNGCNMGNMGNGGNCAYMGNGTNATGGTNGTNA 796

Qy    514 CTTCAGCGGCACGCGTATATCTGGACGGGAATCCACTGTCTGTACGCACTCGTGACAGGC 573
           : |        | | :||:   || |  :    :|    || :| :  :|  |:| :
Db    797 THCCNMGNTAYGGNGARTAYGCNGARGCNMGNGAYYTNGGNGTNMGNMGNMGNTAYAARG 856

Qy    574 TCTGCGGACATCATTGGCCATTCAGGCATCAGGATACACTTCGATATGGCGGGGCCTTCC 633
          |    || ||  |:::   : | |   |:    |:::   :  : |:|| |  | :||:
Db    857 TNGCNGGNCARGAYWSNGARGTNACNTAYTTYCAYWSNTAYATHGAYGGNGTNGAYTTYG 916

Qy    634 GTCCCCGGGAAGCCCGGGCACTGCACC 660
            :  ||:|| |   |  | :  |
Db    917 TNTTYGTNGARGCNCCNCCNTTYMGNC 943
```

RESULT 14
V66834
ID V66834 standard; DNA; 2380 BP.
AC V66834;
DT 05-JAN-1999 (first entry)
DE Zea mays soluble starch synthase gene SSS56.
KW Zea mays; US yellow-dent corn line; maize; soluble starch synthase;
KW SSS; glycogen biosynthetic pathway; branching enzyme; ss.
OS Zea mays.
PN US5824790-A.
PD 20-OCT-1998.
PF 15-DEC-1995; 572951.
PR 15-DEC-1995; US-572951.
PR 21-JUN-1994; US-263921.
PR 29-NOV-1994; US-346602.
PA (ZENE) ZENECA LTD.
PI Guan H, Keeling PL, Knight ME;
DR WPI; 98-582626/49.
PT Isolated nucleic acid molecule, used to produce transgenic plants -
PT comprises nucleotide sequence encoding polypeptide having soluble
PT starch synthase activity, where polypeptide is encoded by maize gene
PS Claim 3; Column 31-33; 29pp; English.
CC The present sequence represents an isolated nucleic acid molecule which
CC has been isolated and comprises a nucleotide sequence encoding a
CC polypeptide having soluble starch synthase (SSS) activity, where the
CC polypeptide is encoded by a maize gene. The isolated nucleic acid
CC molecule can be used to produce transgenic plants with altered starch
CC production. The transgenic plants produced using the nucleic acid
CC molecule have an enhanced ability to produce structurally-altered starch.
SQ Sequence 2380 BP; 333 A; 266 C; 518 G; 272 T;

```
  Query Match             2.1%;  Score 32.4;  DB 1;  Length 2380;
  Best Local Similarity   28.5%;  Pred. No. 8.3;
  Matches   59;  Conservative   31;  Mismatches  117;  Indels    0;  Gaps    0;

Qy    454 AGGAGCTGGATGTCTCTCATAACCAGCTGACCCGCCTGCCGCAAAGCCTCACGGGTCTGT 513
          | |   | |  |  :  |  ||    :|  | :| :     || :|  | | ||   |
Db    737 AYGTNGTNGGNGCNYTNCCNAARGCNYTNGCNMGNMGNGGNCAYMGNGTNATGGTNGTNA 796

Qy    514 CTTCAGCGGCACGCGTATATCTGGACGGGAATCCACTGTCTGTACGCACTCGTGACAGGC 573
           : |        | | :||:   || |  :    :|    || :| :  :|  |:| :
Db    797 THCCNMGNTAYGGNGARTAYGCNGARGCNMGNGAYYTNGGNGTNMGNMGNMGNTAYAARG 856

Qy    574 TCTGCGGACATCATTGGCCATTCAGGCATCAGGATACACTTCGATATGGCGGGGCCTTCC 633
          |    || ||  |:::   : | |   |:    |:::   :  : |:|| |  | :||:
Db    857 TNGCNGGNCARGAYWSNGARGTNACNTAYTTYCAYWSNTAYATHGAYGGNGTNGAYTTYG 916

Qy    634 GTCCCCGGGAAGCCCGGGCACTGCACC 660
            :  ||:|| |   |  | :  |
Db    917 TNTTYGTNGARGCNCCNCCNTTYMGNC 943
```

RESULT 15
Q87790/c
ID Q87790 standard; DNA; 2945 BP.
AC Q87790;
DT 25-OCT-1995 (first entry)
DE S. lavendulae mcr locus.
KW Mitomycin-C resistance; mcr locus; DNA alkylating agent;
KW DNA cleavage agent; drug resistance; ds.
OS Streptomyces lavendulae.

```
FH  Key             Location/Qualifiers
FT  -35_signal      91. .96
FT                  /*tag= a
FT                  /note= "potential -35 region for translational
FT                  start point for promoter P1"
FT  -10_signal      118. .123
FT                  /*tag= b
FT                  /note= "potential -10 region for translational
FT                  start point for promoter P2"
FT  misc_signal     131
FT                  /*tag= c
FT                  /label= P1
FT                  /note= "transcriptional start point 1"
FT  cds             131. .1474
FT                  /*tag= d
FT                  /label= MCRA1
FT  -35_signal      131. .142
FT                  /*tag= e
FT                  /note= "potential -35 region for translational
FT                  start point 2"
FT  -10_signal      156. .162
FT                  /*tag= f
```

```
FT                  /note= "potential -10 region of translational
FT                  start point 2"
FT   misc_signal    170
FT                  /*tag=  g
FT                  /label= P2
FT                  /note= "transcriptional start point 2"
FT   cds            299. .1474
FT                  /*tag=  h
FT                  /label= MCRA1
FT                  /note= "shortened mcra1 transcript"
FT   rbs            284. .290
FT                  /*tag=  i
FT                  /note= "potential RBS for shortened mcra1
FT                  transcript"
FT   cds            1527. .1841
FT                  /*tag=  j
FT                  /label= MCRA2
FT   cds            1962. .2803
FT                  /*tag=  k
FT                  /label= ORF3
PN   WO9509926-A.
PD   13-APR-1995.
PF   06-OCT-1994; U11279.
PR   07-OCT-1993; US-133963.
PA   (MINU ) UNIV MINNESOTA.
PI   August PR,  Flickinger MC,  Sherman DH;
DR   WPI; 95-155267/20.
DR   P-PSDB; R72379; R72380; R72381.
PT   Genes encoding resistance to DNA alkylating and cleavage agents
PT   - used to combat drug resistance of tumour cells or as potential
PT   anticancer agents
PS   Disclosure; Fig.2; 111pp; English.
CC   A 3.2 kb fragment from pDHS3003 including the mcr locus of S.
CC   lavendulae B619 was cloned into pUC119 and sequenced.  The locus
CC   included 3 coding sequences, mcrA, mcrB and mcrORF3.  Both mcrA and
CC   B are needed for high level resistance to mitomycin-C.
SQ   Sequence   2945 BP;    451 A;    1029 C;    1022 G;    443 T;


  Query Match             2.1%;  Score 31.8;  DB 1;  Length 2945;
  Best Local Similarity   56.1%;  Pred. No. 14;
  Matches   60;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy     632 CCGTCCCCGGGAAGCCCGGGCACTGCACCTGGCGGTCGCTGACTGGCTGACGTCTGCACG 691
           ||  ||||| |   ||   |  | |  | |||| ||||   |  ||   ||  ||| |
Db    1833 CCTCCCCCGCGGTTCCACCGGCCCGGTCATGGCCTTCGCCTGCCAGCAACCGCTTGCGCT 1774

Qy     692 GGAGGGGGAAGCGGCCCAGGCAGACAGATGGCAGGCGTTCGGACTGG 738
             ||| | || | | |  |  |  ||  |||||| ||||| ||||
Db    1773 CCGGGGTGTAGTGCCACCAGGGGCGAGCGGGCAGGCCTTCGGCCTGG 1727


Search completed: October  5, 1999, 18:00:59
Job time: 11501 sec
```

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 1999, 18:04:03 ; Search time 2414.69 Seconds
(without alignments)
2025.650 Million cell updates/sec

Title: US-09-068-804-13
Perfect score: 1538
Sequence: 1 ATGCGTGATTGCCTGAATAA.....CCCGGTGGCGTCTTAACTGA 1538

Scoring table: IDENTITY_NUC

Searched: 679419 seqs, 1590154680 residues

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_st:*
14: gb_sts:*
15: gb_sy:*
16: gb_un:*
17: gb_vi:*
18: em_fun:*
19: em_htg:*
20: em_hum1:*
21: em_hum2:*
22: em_in:*
23: em_om:*
24: em_or:*
25: em_ov:*
26: em_pat:*
27: em_ph:*
28: em_pl:*
29: em_ro:*
30: em_sts:*
31: em_sy:*
32: em_un:*
33: em_vi:*
34: gb_htg1:*
35: gb_htg2:*
36: gb_in1:*
37: gb_in2:*
38: em_ba1:*
39: em_ba2:*
40: em_hum3:*
41: em_hum4:*
42: gb_pr4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	268.6	17.5	1992	1	SHFIPAHZ	M76445 Shigella fl
2	223.4	14.5	2900	1	SHFIPAH	M32063 S.flexner i
3	217.4	14.1	1822	2	AF047365	AF047365 Shigella
4	216.8	14.1	1312	1	SHFIPAHX	M76443 Shigella fl
5	216.8	14.1	1375	1	SHFIPAHY	M76444 Shigella fl
6	142.6	9.3	14263	2	AE000074	AE000074 Rhizobium
c 7	45.4	3.0	7218	5	I66494	I66494 Sequence 14
8	42.6	2.8	30683	2	AF027770	AF027770 Mycobacte
9	42.6	2.8	15223	2	AF034152	AF034152 Mycobacte
10	41.6	2.7	2935	1	BACLEVRA	M60105 Bacillus su
11	41.6	2.7	23911	1	BS233DEG	X92868 B.subtilis
c 12	41.6	2.7	213420	1	BSUB0014	Z99117 Bacillus su
c 13	40.6	2.6	1908	9	GORINVOLUA	M23603 Gorilla gor
14	40	2.6	39739	1	SC10A7	AL078618 Streptomy
15	39.8	2.6	1397	1	YEPYOPM	M25810 Y.pestis ou
16	39.8	2.6	70504	2	AF053946	AF053946 Yersinia
17	39.8	2.6	70559	2	AF074612	AF074612 Yersinia
c 18	39.6	2.6	156641	11	HUAC002544	AC002544 Homo sapi
c 19	39.6	2.6	150296	11	HUU91326	U91326 Human Chrom
20	39.4	2.6	6471	9	AB007914	AB007914 Homo sapi
c 21	38	2.5	12581	1	SCE36	AL049763 Streptomy
22	37.6	2.4	1766	12	MUSF8A	M83118 Mus musculu
c 23	37.4	2.4	1818	9	GORINVOLUB	M23604 Gorilla gor
c 24	37.4	2.4	157093	34	AC006916	AC006916 Caenorhab
c 25	37.2	2.4	26440	1	SCE15	AL049707 Streptomy
c 26	37.2	2.4	1150	5	I40981	I40981 Sequence 10
27	36.8	2.4	1018	3	OAKRTAPB	X73435 O.aries KRT
c 28	36.6	2.4	11535	2	AE000949	AE000949 Archaeogl
c 29	36.6	2.4	2224	11	HSU80736	U80736 Homo sapien
30	36.6	2.4	2479	12	RNU87960	U87960 Rattus norv
31	36.4	2.4	1355	12	MMU20553	U20553 Mus musculu
c 32	36.4	2.4	146126	34	HSDJ977B1	AL050318 Homo sapi
c 33	36.4	2.4	286845	35	AC006449	AC006449 Homo sapi
c 34	36.2	2.4	2872	1	SCSCOF	X92686 S.coelicolo
c 35	36.2	2.4	2479	12	RNU87960	U87960 Rattus norv
c 36	36	2.3	38404	1	SC2G5	AL035478 Streptomy
c 37	35.8	2.3	93353	7	ATF10M6	AL021811 Arabidops
38	35.6	2.3	2490	1	AB019394	AB019394 Streptomy
c 39	35.6	2.3	2481	1	ECOPRFC	D17724 Escherichia
c 40	35.6	2.3	338534	1	ECOUW93	U14003 Escherichia
c 41	35.6	2.3	2324	1	ECPRFC	Z26313 E.coli prfC
c 42	35.6	2.3	12359	2	AE000507	AE000507 Escherich
43	35.6	2.3	9296	3	OCILIPBP	Z54345 O.cuniculus
44	35.6	2.3	2155	5	A26494	A26494 Human IDUA
45	35.6	2.3	1758	5	A26496	A26496 Human IDUA

ALIGNMENTS

RESULT 1
SHFIPAHZ
LOCUS SHFIPAHZ 1992 bp DNA BCT 23-OCT-1991
DEFINITION Shigella flexneri ipaH gene, 5'end.
ACCESSION M76445
NID g152751
VERSION M76445.1 GI:152751
KEYWORDS LRG-like protein; ipaH gene.
SOURCE Shigella flexneri (strain 5) DNA.
ORGANISM Shigella flexneri
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Shigella.
REFERENCE 1 (bases 1 to 1992)
AUTHORS Hartman,A.B., Venkatesan,M., Oaks,E.V. and Buysse,J.M.
TITLE Sequence and molecular characterization of a multicopy invasion plasmid antigen gene, ipaH, of Shigella flexneri
JOURNAL J. Bacteriol. 172, 1905-1915 (1990)
MEDLINE 90202708
REFERENCE 2 (bases 1 to 1992)
AUTHORS Venkatesan,M.M., Buysse,J.M. and Hartman,A.B.
TITLE Sequence variation in two ipaH genes of Shigella flexneri 5 and homology to the LRG-like family of proteins
JOURNAL Mol. Microbiol. 5, 2435-2446 (1991)
MEDLINE 92167809

```
FEATURES             Location/Qualifiers
     source          1. .1992
                     /organism="Shigella flexneri"
                     /strain="5"
                     /db_xref="taxon:623"
     gene            1. .1725
                     /gene="ipaH"
     CDS             1. .1725
                     /gene="ipaH"
                     /codon_start=1
                     /transl_table=11
                     /product="antigen"
                     /protein_id="AAA26528.1"
                     /db_xref="PID:g152752"
                     /db_xref="GI:152752"
                     /translation="MKPINNHSFFRSLCGLSCISRLSVEEQCTRDYHRIWDDWAREGT
                     TTENRIQAVRLLKICLDTREPVLNLSLLKLRSLPPLPLHIRELNISNNELISLPENSP
                     LLTELHVNGNNLNILPTLPSQLIKLNISFNRNLSCLPSLPPYLQSLSARFNSLETLPE
                     LPSTLTILRIEGNRLTVLPELPHRLQELFVSGNRLQELPEFPQSLKYLKVGENQLRRL
                     SRLPQELLALDVSNNLLTSLPENIITLPICTNVNISGNPLSTRVLQSLQRLTSSPDYH
                     GPQIYFSMSDGQQNTLHRPLADAVTAWFPENKQSDVSQIWHAFEHEEHANTFSAFLDR
                     LSDTVSARNTSGFREQVAAWLEKLSASAELRQQSFAVAADATESCEDRVALTWNNLRK
                     TLLVHQASEGLFDNDTGALLSLGREMFRLEILEDIARDKVRTLHFVDEIEVYLAFQTM
                     LAEKLQLSTAVKEMRFYGVSGVTANDLRTAEAMVRSREENEFTDWFSLWGPWHAVLKR
                     TEADRWAQAEEQKYEMLENEYSQRVADRLKASGLSGDADAQREAGAQVMRETEQQIYR
                     QLTDEVLA"
BASE COUNT      529 a    500 c    464 g    499 t
ORIGIN
```

Query Match 17.5%; Score 268.6; DB 1; Length 1992;
Best Local Similarity 50.5%; Pred. No. 1.8e-62;
Matches 703; Conservative 0; Mismatches 684; Indels 4; Gaps 2;

```
Qy      80 CACCGCATATTACAACACTGGTTATTCCTGATAATAATCTGACCAGCCTGCCGGAGTTGC 139
Db     302 CTCCGCTTTTGACAGAACTTCATGTAAATGGTAACAACTTGAATATACTCCCGACACTTC 361

Qy     140 CGGAAGGACTACGGGAGCTGGAGGTCTCTGGTAACCTACAACTGACCAGCCTGCCATCGC 199
Db     362 CATCTCAACTGATTAAGCTTAATATTTCATTCAATCGAAATTTGTCATGTCTGCCATCAT 421

Qy     200 TGCCGCAGGGACTACAGAAGCTGTGGGCCTATAATAATTGGCTGGCCAGCCTGCCGACGT 259
Db     422 TACCACCATATTTACAATCACTCTCGGCACGTTTTAATAGTCTGGAGACGTTACCAGAGC 481

Qy     260 TGCCGCCAGGACTAGGGGATCTGGCGGTCTCTAATAACCAGCTGACCAGCCTGCCGGAGA 319
Db     482 TTCCATCAACGCTAACAATATTACGTATTGAAGGTAATCGCCTTACTGTCTTGCCTGAAT 541

Qy     320 TGCCGCCAGCACTACGGGAGCTGAGGGTCTCTGGTAACA-ACCTGACCAGCTGCGCGCGC 378
Db     542 TGCCTCATAGACTACAAGAACTCTTTGTTTCCGGCAACAGACTACAGGAACTACCAGAAT 601

Qy     379 TGCCGTCAGGACTACAGAAGCTGTGGGCCTATAATAATCGGCTGACCAGCCTGCCGGAGA 438
Db     602 TTCCTCAGAGCTTAAAATATTTGAAGGTAGGTGAAAATCAACTACGCAGATTATCCAGAT 661

Qy     439 TGTCGCCAGGACTACAGGAGCTGGATGTCTCTCATAACCAGCTGACCCGCCTGCCGCAAA 498
Db     662 TACCGCAAGAACTATTGGCTCTGGATGTTTCCAATAACCTACTAACTTCATTACCCGAAA 721

Qy     499 GCCTCACGGGTCTGTCTTCAGCGGCACGCGTATATCTGGACGGGAATCCACTGTCTGTAC 558
Db     722 ATATAATCACATTGCCCATTTGTACGAATGTTAACATTTCAGGGAATCCATTGTCGACTC 781

Qy     559 GCACTCGTGA---CAGGCTCTGCGGACATCATTGGCCATTCAGGCATCAGGATACACTTC 615
Db     782 GCGTTCTGCAATCCCTGCAAAGATTAACCTCTTCGCCGGACTACCACGGCCCGCAGATTT 841

Qy     616 GATATGGCGGGGCCTTCCGTCCCCGGGAAGCCCGGGCACTGCACCTGGCGGTCGCTGACT 675
Db     842 ACTTCTCCATGAGTGACGGACAACAGAATACACTCCATCGCCCCCTGGCTGATGCCGTGA 901

Qy     676 GGCTGACGTCTGCACGGGAGGGGGAAGCGGCCCAGGCAGACAGATGGCAGGCGTTCGGAC 735
Db     902 CAGCATGGTTCCCGGAAAACAAACAATCTGATGTATCACAGATATGGCATGCTTTTGAAC 961

Qy     736 TGGAAGATAACGCCGCCGCCTTCAGCCTGGTCCTGGACAGACTGCGTGAGACGGAAAACT 795
Db     962 ATGAAGAGCATGCCAACACCTTTTCCGCGTTCCTTGACCGCCTTTCCGATACCGTCTCTG 1021

Qy     796 TCAAAAAAGACGCGGGCTTTAAGGCACAGATATCATCCTGGCTGACACAACTGGCTGAAG 855
Db    1022 CACGCAATACCTCCGGATTCCGTGAACAGGTCGCTGCATGGCTGGAAAAACTCAGTGCCT 1081

Qy     856 ATGCTGCGCTGAGAGCAAAAACCTTTGCCATGGCAACAGAGGCAACATCAACCTGCGAGG 915
Db    1082 CTGCGGAGCTTCGACAGCAGTCTTTCGCTGTTGCTGCTGATGCCACTGAGAGCTGTGAGG 1141

Qy     916 ACCGGGTCACACATGCCCTGCACCAGATGAATAACGTACAACTGGTACATAATGCAGAAA 975
Db    1142 ACCGTGTCGCGCTCACATGGAACAATCTCCGGAAAACCCTCCTGGTCCATCAGGCATCAG 1201

Qy     976 AAGGGGAATACGACAACAATCTCCAGGGGCTGGTTTCCACGGGGCGTGAGATGTTCCGCC 1035
Db    1202 AAGGCCTTTTCGATAATGATACCGGCGCTCTGCTCTCCCTGGGCAGGGAAATGTTCCGCC 1261

Qy    1036 TGGCAACACTGGAACAGATTGCCCGGGAAAAAGCCGGAACACTGGCTTTAGTCGATGACG 1095
Db    1262 TCGAAATTCTGGAGGACATTGCCCGGGATAAAGTCAGAACTCTCCATTTTGTGGATGAGA 1321

Qy    1096 TTGAGGTCTATCTGGCGTTCCAGAATAAGCTGAAGGAATCACTTGAGCTGACCAGCGTGA 1155
Db    1322 TAGAAGTCTACCTGGCCTTCCAGACCATGCTCGCAGAGAAACTTCAGCTCTCCACTGCCG 1381

Qy    1156 CGTCAGAAATGCGTTTCTTTGACGTTTCCGGCGTGACGGTTTCAGACCTTCAGGCTGCGG 1215
Db    1382 TGAAGGAAATGCGTTTCTATGGCGTGTCGGGAGTGACAGCAAATGACCTCCGCACTGCCG 1441

Qy    1216 ACGTTCAGGTGAAAACCGCTGAAAACAGCGGGTTCAGTAAATGGATACTGCAGTGGGGGC 1275
Db    1442 AAGCTATGGTCAGAAGCCGTGAAGAGAATGAATTTACGGACTGGTTCTCCCTCTGGGGAC 1501

Qy    1276 CGTTACACAGCGTGCTGGAACGCAAAGTGCCGGAACGCTTTAACGCGCTTCGTGAAAAGC 1335
Db    1502 CATGGCATGCTGTACTGAAGCGTACGGAAGCTGACCGCTGGGCGCAGGCAGAAGAGCAGA 1561

Qy    1336 AAATATCGGATTATGAAGACACGTACCGGAAGCTGTATGACGAAGTGCTGAAATCGTCCG 1395
Db    1562 AGTATGAGATGCTGGAGAATGAGTACTCTCAGAGGGTGGCTGACCGGCTGAAAGCATCAG 1621

Qy    1396 GGCTGGTCGACGATACCGATGCAGAACGTACTATCGGAGTAAGTGCGATGGATAGTGCGA 1455
Db    1622 GTCTGAGCGGTGATGCGGATGCGCAGAGGGAAGCCGGTGCACAGGTGATGCGTGAGACTG 1681

Qy    1456 AAAAAGAATTT 1466
Db    1682 AACAGCAGATT 1692
```

```
RESULT   2
SHFIPAH
LOCUS       SHFIPAH      2900 bp    DNA             BCT       23-JAN-1991
DEFINITION  S.flexner invasion plasmid antigen H (ipaH) gene, complete cds.
ACCESSION   M32063
NID         g152746
VERSION     M32063.1  GI:152746
KEYWORDS    invasion plasmid antigen.
SOURCE      S.flexner (strain M90T-W; serotype 5) DNA, clone pWR390.
  ORGANISM  Shigella flexneri
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Shigella.
REFERENCE   1  (bases 1 to 2900)
  AUTHORS   Hartman,A.B., Venkatesan,M., Oaks,E.V. and Buysse,J.M.
  TITLE     Sequence and molecular characterization of a multicopy invasion
```

```
            plasmid antigen gene, ipaH, of Shigella flexneri
  JOURNAL   J. Bacteriol. 172, 1905-1915 (1990)
  MEDLINE   90202708
COMMENT     Draft entry and computer-readable sequence [1] kindly submitted by
            A.B.Hartman, 12-FEB-1990.
FEATURES             Location/Qualifiers
     source          1. .2900
                     /organism="Shigella flexneri"
                     /db_xref="taxon:623"
     -35_signal      139. .144
     -10_signal      162. .167
     CDS             251. .1849
                     /note="invasion plasmid antigen H"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA26526.1"
                     /db_xref="PID:g152747"
                     /db_xref="GI:152747"
                     /translation="MGKELFSREERGIAFNRLSQCFQNQEAVLNLSDLNLTSLPELPK
                     HISALIVENNKLTSLPKLPAFLKELNADNNRLSVIPELPESLTTLSVRSNQLENLPVL
                     PNHLTSLFVENNRLYNLPALPEKLKFLHVYYNRLTTLPDLPDKLEILCAQRNNLVTFP
                     QFSDRNNIRQKEYYFHFNQITTLPESFSQLDSSYRINISGNPLSTRVLQSLQRLTSSP
                     DYHGPQIYFSMSDGQQNTLHRPLADAVTAWFPENKQSDVSQIWHAFEHEEHANTFSAF
                     LDRLSDTVSARNTSGFREQVAAWLEKLSASAELRQQSFAVAADATESCEDRVALTWNN
                     LRKTLLVHQASEGLFDNDTGALLSLGREMFRLEILEDIARDKVRTLHFVDEIEVYLAF
                     QTMLAEKLQLSTAVKEMRFYGVSGVTANDLRTAEAMVRSREENEFTDWFSLWGPWHAV
                     LKRTEADRWAQAEEQKYEMLENEYSQRVADRLKASGLSGDADAQREAGAQVMRETEQQ
                     IYRQLTDEVLALRLSENGSRLHHS"
     -35_signal      2111. .2116
     -10_signal      2132. .2137
     CDS             2277. .>2900
                     /note="ORF3"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA26527.1"
                     /db_xref="PID:g152748"
                     /db_xref="GI:152748"
                     /translation="MKPINNHSFFRSLCGLSCISRLSVEEQCTRDYHRIWDDWAREGT
                     TTENRIQAVRLLKICLDTREPVLNLSLLKLRSLPPLPLHIRELNISNNELISLPENSP
                     LLTELHVNGNNLNILPTLPSQLIKLNISFNRNLSCLPSLPPYLQSLSARFNSLETLPE
                     LPSTLTILRIEGNRLTVLPELPHRLQELFVSGNRLQELPEFPQSLKYL"
BASE COUNT      868 a    657 c    578 g    797 t
ORIGIN
```

```
  Query Match             14.5%;  Score 223.4;  DB 1;  Length 2900;
  Best Local Similarity   49.5%;  Pred. No. 3.2e-50;
  Matches  697;  Conservative    0;  Mismatches  691;  Indels   19;  Gaps    4;

Qy      76 TTACCACCGCATATTACAACACTGGTTATTCCTGATAATAATCTGACCAGCCTGCCGGAG 135
           ||||||  |||||||  |  |  || || |      |||||||  | ||     ||||  ||
Db     374 TTACCAAAGCATATTTCTGCTTTGATTGTAGAAAATAATAAATTAACATCATTGCCAAAG 433

Qy     136 TTGCCGGAAGGACTACGGGAGCTGGAGGTCTCTGGTAACCTACAACTGACCAGCCTGCCA 195
           ||||  |  |    ||    || ||  | |   ||| |||        ||  |      | |||
Db     434 CTGCCTGCATTTCTTAAAGAACTTAATG---CTGATAATAACAGGCTTTCTGTGATACCA 490

Qy     196 TCGCTGCCGCAGGGACTACAGAAGCTGTGGGCCTATAATAATTGGCTGGCCAGCCTGCCG 255
           || ||  ||  | ||    |  | | |    | ||||   ||||  | ||| ||
Db     491 GAACTTCCTGAGTCATTAACAACTTTAAGTGTTCGTTCTAATCAACTGGAAAACCTTCCT 550

Qy     256 ACGTTGCCGCCAGGACTAGGGGATCTGGCGGTCTCTAATAACCAGCTGACCAGCCTGCCG 315
           |||||         ||        |    ||    ||||||  |||    | | | |||
Db     551 GTTTTGCCAAACCATTTAACATCATTATTTGTTGAAAATAACAGGCTATATAACTTACCG 610

Qy     316 GAGATGCCGCCAGCACTACGGGAGCTGAGGGTCTCTGGTAACAACCTGACCAGCTG-CGC 374
           |   | ||   | | |       |    || | |  |||||  |||||  |   |  | |
Db     611 GCTCTTCCCGAAAAATTGAAATTTTTACATGTTTATTATAACAGGCTGACAACATTACCC 670

Qy     375 GCGCTGCCGTCAGGACTACAGAAGCTGTGGGCCTATAATAATCGGCTGACCAGCCTGCCG 434
           |   | |||     |||  | |  || || ||   |     |||   |||   |   | ||
Db     671 GACTTACCGGATAAACTGGAAATTCTCTGTGCTCAGCGCAATAATCTGGTTACTTTTCCT 730

Qy     435 GAGATGTCG----------CCAGGACTACAGGAGCTGGATGTCTCTCATAACCAGCTG 482
           |  | ||                  ||| | ||||    || |   |  ||| ||| |
Db     731 CAATTTTCTGATAGAAACAATATCAGACAAAAGGAATATTATTTTCATTTTAATCAGATA 790

Qy     483 ACCCGCCTGCCGCAAAGCCTCACGGGTCTGTCTTCAGCGGCACGCGTATATCTGGACGGG 542
           |||   || ||| | ||  |  |    |  ||||       |  | || |  |||
Db     791 ACCACTCTTCCGGAGAGTTTTTCACAATTAGATTCAAGTTACAGGATTAATATTTCAGGG 850

Qy     543 AATCCACTGTCTGTACGCACTCGTGA---CAGGCTCTGCGGACATCATTGGCCATTCAGG 599
           |||||| ||||    |||  ||   |   |  ||  |   |      || |||   |
Db     851 AATCCATTGTCGACTCGCGTTCTGCAATCCCTGCAAAGATTAACCTCTTCGCCGGACTAC 910

Qy     600 CATCAGGATACACTTCGATATGGCGGGGCCTTCCGTCCCCGGGAAGCCCGGGCACTGCAC 659
           ||          ||   |  |  |  | | | | | | |  |     | | |
Db     911 CACGGCCCGCAGATTTACTTCTCCATGAGTGACGGACAACAGAATACACTCCATCGCCCC 970

Qy     660 CTGGCGGTCGCTGACTGGCTGACGTCTGCACGGGAGGGGGAAGCGGCCCAGGCAGACAGA 719
           |||||  |  || |          ||   |      |    || | |       || | | |
Db     971 CTGGCTGATGCCGTGACAGCATGGTTCCCGGAAAACAAACAATCTGATGTATCACAGATA 1030

Qy     720 TGGCAGGCGTTCGGACTGGAAGATAACGCCGCCGCCTTCAGCCTGGTCCTGGACAGACTG 779
           |||||  || ||  | ||  |||||  | |||  | ||||   |  | |||| ||| | ||
Db    1031 TGGCATGCTTTTGAACATGAAGAGCATGCCAACACCTTTTCCGCGTTCCTTGACCGCCTT 1090

Qy     780 CGTGAGACGGAAAACTTCAAAAAAGACGCGGGCTTTAAGGCACAGATATCATCCTGGCTG 839
           | || | |          ||   | | || ||    | |||| |  |  | ||||||
Db    1091 TCCGATACCGTCTCTGCACGCAATACCTCCGGATTCCGTGAACAGGTCGCTGCATGGCTG 1150

Qy     840 ACACAACTGGCTGAAGATGCTGCGCTGAGAGCAAAAACCTTTGCCATGGCAACAGAGGCA 899
           | ||||   ||   ||| | |||  ||    |  | || ||  | ||  | || ||
Db    1151 GAAAAACTCAGTGCCTCTGCGGAGCTTCGACAGCAGTCTTTCGCTGTTGCTGCTGATGCC 1210

Qy     900 ACATCAACCTGCGAGGACCGGGTCACACATGCCCTGCACCAGATGAATAACGTACAACTG 959
           ||    | ||| |||||||| ||| | |   |  | || |  |   ||   | |||
Db    1211 ACTGAGAGCTGTGAGGACCGTGTCGCGCTCACATGGAACAATCTCCGGAAAACCCTCCTG 1270

Qy     960 GTACATAATGCAGAAAAAGGGGAATACGACAACAATCTCCAGGGGCTGGTTTCCACGGGG 1019
           || ||| | |||  | ||||    | ||| ||  ||  |   |  ||| | |||  |||
Db    1271 GTCCATCAGGCATCAGAAGGCCTTTTCGATAATGATACCGGCGCTCTGCTCTCCCTGGGC 1330

Qy    1020 CGTGAGATGTTCCGCCTGGCAACACTGGAACAGATTGCCCGGGAAAAAGCCGGAACACTG 1079
           | || |||||||||| | ||  |||||  | |||||||||| |||| | |||| ||
Db    1331 AGGGAAATGTTCCGCCTCGAAATTCTGGAGGACATTGCCCGGGATAAAGTCAGAACTCTC 1390

Qy    1080 GCTTTAGTCGATGACGTTGAGGTCTATCTGGCGTTCCAGAATAAGCTGAAGGAATCACTT 1139
           |||  || |||||  | || |||||  |||||  |||||||  | |||    ||  ||||
Db    1391 CATTTTGTGGATGAGATAGAAGTCTACCTGGCCTTCCAGACCATGCTCGCAGAGAAACTT 1450

Qy    1140 GAGCTGACCAGCGTGACGTCAGAAATGCGTTTCTTTGACGTTTCCGGCGTGACGGTTTCA 1199
           ||||  |||  |   |  ||||||||||||| || ||| || || |||||  |
Db    1451 CAGCTCTCCACTGCCGTGAAGGAAATGCGTTTCTATGGCGTGTCGGGAGTGACAGCAAAT 1510

Qy    1200 GACCTTCAGGCTGCGGACGTTCAGGTGAAAACCGCTGAAAACAGCGGGTTCAGTAAATGG 1259
           ||||| |   |||| || | |  ||| | || |  |||  | |  |  || |   | |||
Db    1511 GACCTCCGCACTGCCGAAGCTATGGTCAGAAGCCGTGAAGAGAATGAATTTACGGACTGG 1570

Qy    1260 ATACTGCAGTGGGGGCCGTTACACAGCGTGCTGGAACGCAAAGTGCCGGAACGCTTTAAC 1319
           |    |  |||| || |  ||    || ||| | || |  |  |  | || ||||
Db    1571 TTCTCCCTCTGGGGACCATGGCATGCTGTACTGAAGCGTACGGAAGCTGACCGCTGGGCG 1630

Qy    1320 GCGCTTCGTGAAAAGCAAATATCGGATTATGAAGACACGTACCGGAAGCTGTATGACGAA 1379
           |     ||  || |       ||  |  ||||    ||  |   |  ||
Db    1631 CAGGCAGAAGAGCAGAAGTATGAGATGCTGGAGAATGAGTACTCTCAGAGGGTGGCTGAC 1690

Qy    1380 GTGCTGAAATCGTCCGGGCTGGTCGACGATACCGATGCAGAACGTACTATCGGAGTAAGT 1439
           |||||||  | || || ||| ||  ||| | ||||  |  |  |       ||| | |
Db    1691 CGGCTGAAAGCATCAGGTCTGAGCGGTGATGCGGATGCGCAGAGGGAAGCCGGTGCACAG 1750

Qy    1440 GCGATGGATAGTGCGAAAAAAGAATTT 1466
           | ||||  |   |  || |  |  ||
Db    1751 GTGATGCGTGAGACTGAACAGCAGATT 1777
```

```
RESULT   3
AF047365
LOCUS       AF047365     1822 bp    DNA             BCT       19-JUN-1998
DEFINITION  Shigella flexneri plasmid pWR100 IpaH9.8 (ipaH9.8) gene, complete
            cds.
ACCESSION   AF047365
NID         g3237376
VERSION     AF047365.1  GI:3237376
KEYWORDS    .
SOURCE      Shigella flexneri.
  ORGANISM  Shigella flexneri
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Shigella.
REFERENCE   1  (bases 1 to 1822)
  AUTHORS   Demers,B., Sansonetti,P.J. and Parsot,C.
  TITLE     Induction of type III secretion in shigella flexneri is associated
            with differential control of transcription of genes encoding
            secreted proteins
  JOURNAL   EMBO J. 17 (10), 2894-2903 (1998)
  MEDLINE   98250661
REFERENCE   2  (bases 1 to 1822)
  AUTHORS   Demers,B., Sansonetti,P.J. and Parsot,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-FEB-1998) Molecular Pathogenic Microbiology, Pasteur
            Institute, 28 rue du Dr. Roux, Paris 75724, France
FEATURES             Location/Qualifiers
     source          1. .1822
                     /organism="Shigella flexneri"
                     /strain="M90T"
                     /db_xref="taxon:623"
                     /plasmid="pWR100"
                     /note="virulence plasmid"
     gene            160. .1797
                     /gene="ipaH9.8"
     CDS             160. .1797
                     /gene="ipaH9.8"
                     /note="secreted by the type III secretion pathway encoded
                     by the virulence plasmid"
                     /codon_start=1
                     /transl_table=11
                     /product="IpaH9.8"
                     /protein_id="AAC23714.1"
                     /db_xref="PID:g3237377"
                     /db_xref="GI:3237377"
                     /translation="MLPINNNFSLPQNSFYNTISGTYADYFSAWDKWEKQALPGEERD
                     EAVSRLKECLINNSDELRLDRLNLSSLPDNLPAQITLLNVSYNQLTNLPELPVTLKKL
                     YSASNKLSELPVLPPALESLQVQHNELENLPALPDSLLTMNISYNEIVSLPSLPQALK
                     NLRATRNFLTELPAFSEGNNPVVREYFFDRNQISHIPESILNLRNECSIHISDNPLSS
                     HALQALQRLTSSPDYHGPRIYFSMSDGQQNTLHRPLADAVTAWFPENKQSDVSQIWHA
                     FEHEEHANTFSAFLDRLSDTVSARNTSGFREQVAAWLEKLSASAELRQQSFAVAADAT
                     ESCEDRVALTWNNLRKTLLVHQASEGLFDNDTGALLSLGREMFRLEILEDIARDKVRT
                     LHFVDEIEVYLAFQTMLAEKLQLSTAVKEMRFYGVSGVTANDLRTAEAMVRSREENEF
                     TDWFSLWGPWHAVLKRTEADRWAQAEEQKYEMLENEYPQRVADRLKASGLSGDADAER
                     EAGAQVMRETEQQIYRQLTDEVLALRLSENGSQLHHS"
BASE COUNT      501 a    457 c    396 g    468 t
ORIGIN
```

```
  Query Match             14.1%;  Score 217.4;  DB 2;  Length 1822;
  Best Local Similarity   55.5%;  Pred. No. 1.3e-48;
  Matches  419;  Conservative    0;  Mismatches  336;  Indels    0;  Gaps    0;

Qy     712 CAGACAGATGGCAGGCGTTCGGACTGGAAGATAACGCCGCCGCCTTCAGCCTGGTCCTGG 771
           || | | |||||| || || | ||  |||||  | |||  | ||||    |  | |||| |
Db     971 CACAGATATGGCATGCTTTTGAACATGAAGAGCATGCCAACACCTTTTCCGCGTTCCTTG 1030

Qy     772 ACAGACTGCGTGAGACGGAAAACTTCAAAAAAGACGCGGGCTTTAAGGCACAGATATCAT 831
           || | ||    || || |         ||   | | || ||    | ||||  |  |
Db    1031 ACCGCCTTTCCGATACCGTCTCTGCACGCAATACCTCCGGATTCCGTGAACAGGTCGCTG 1090

Qy     832 CCTGGCTGACACAACTGGCTGAAGATGCTGCGCTGAGAGCAAAAACCTTTGCCATGGCAA 891
           | ||||||  | ||||   ||    ||| | |||  ||    |  | || ||  | ||
Db    1091 CATGGCTGGAAAAACTCAGTGCCTCTGCGGAGCTTCGACAGCAGTCTTTCGCTGTTGCTG 1150

Qy     892 CAGAGGCAACATCAACCTGCGAGGACCGGGTCACACATGCCCTGCACCAGATGAATAACG 951
           | || || ||    | ||| |||||||| ||| | |   |   | || | |    ||
Db    1151 CTGATGCCACTGAGAGCTGTGAGGACCGTGTCGCGCTCACATGGAACAATCTCCGGAAAA 1210

Qy     952 TACAACTGGTACATAATGCAGAAAAAGGGGAATACGACAACAATCTCCAGGGGCTGGTTT 1011
           |  ||||| ||| | |||  | ||||    | ||| ||  ||  |   |  ||| | |
Db    1211 CCCTCCTGGTCCATCAGGCATCAGAAGGCCTTTTCGATAATGATACCGGCGCTCTGCTCT 1270

Qy    1012 CCACGGGGCGTGAGATGTTCCGCCTGGCAACACTGGAACAGATTGCCCGGGAAAAAGCCG 1071
           ||  |||  | || |||||||||||| | ||  |||||  | ||||||||||| |||| |
Db    1271 CCCTGGGCAGGGAAATGTTCCGCCTCGAAATTCTGGAGGATATTGCCCGGGATAAAGTCA 1330

Qy    1072 GAACACTGGCTTTAGTCGATGACGTTGAGGTCTATCTGGCGTTCCAGAATAAGCTGAAGG 1131
           |||| ||   ||| || |||||  | || ||||| ||||| |||||||  | |||    |
Db    1331 GAACTCTCCATTTTGTGGATGAGATAGAAGTCTACCTGGCCTTCCAGACCATGCTCGCAG 1390

Qy    1132 AATCACTTGAGCTGACCAGCGTGACGTCAGAAATGCGTTTCTTTGACGTTTCCGGCGTGA 1191
           |  |||| |||| ||| |   |  ||||||||||||| || ||| || || ||||
Db    1391 AGAAACTTCAGCTCTCCACTGCCGTGAAGGAAATGCGTTTCTATGGCGTGTCGGGAGTGA 1450

Qy    1192 CGGTTTCAGACCTTCAGGCTGCGGACGTTCAGGTGAAAACCGCTGAAAACAGCGGGTTCA 1251
           | |     ||||| |   |||| || |   ||| | || |  |||| | |  |  || |
Db    1451 CAGCAAATGACCTCCGCACTGCCGAAGCCATGGTCAGAAGCCGTGAAGAGAATGAATTTA 1510

Qy    1252 GTAAATGGATACTGCAGTGGGGGCCGTTACACAGCGTGCTGGAACGCAAAGTGCCGGAAC 1311
           | ||| |   |  ||||| || | ||   || ||| | || |  |   | || |
Db    1511 CGGACTGGTTCTCCCTCTGGGGACCATGGCATGCTGTACTGAAGCGTACGGAAGCTGACC 1570

Qy    1312 GCTTTAACGCGCTTCGTGAAAAGCAAATATCGGATTATGAAGACACGTACCGGAAGCTGT 1371
           |||       |      ||  || ||      |      ||  |   |||||  ||  |
Db    1571 GCTGGGCGCAGGCAGAAGAGCAGAAATATGAGATGCTGGAGAATGAGTACCCTCAGAGGG 1630

Qy    1372 ATGACGAAGTGCTGAAATCGTCCGGGCTGGTCGACGATACCGATGCAGAACGTACTATCG 1431
           |  ||   ||||||| | || || |||  ||  ||| | ||||| ||  |       ||
Db    1631 TGGCTGACCGGCTGAAAGCATCAGGTCTGAGCGGTGATGCGGATGCGGAGAGGGAAGCCG 1690

Qy    1432 GAGTAAGTGCGATGGATAGTGCGAAAAAAGAATTT 1466
           | | |   | ||||  |    |  || | |  ||
Db    1691 GTGCACAGGTGATGCGTGAGACTGAACAGCAGATT 1725
```

```
RESULT   4
SHFIPAHX
LOCUS       SHFIPAHX     1312 bp    DNA             BCT       23-OCT-1991
DEFINITION  Shigella flexneri ipaH gene.
ACCESSION   M76443
NID         g152749
VERSION     M76443.1  GI:152749
KEYWORDS    LRG-like protein; ipaH gene.
SOURCE      Shigella flexneri (strain 5) DNA.
  ORGANISM  Shigella flexneri
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Shigella.
REFERENCE   1  (bases 1 to 1312)
  AUTHORS   Venkatesan,M.M., Buysse,J.M. and Hartman,A.B.
  TITLE     Sequence variation in two ipaH genes of Shigella flexneri 5 and
            homology to the LRG-like family of proteins
  JOURNAL   Mol. Microbiol. 5, 2435-2446 (1991)
  MEDLINE   92167809
FEATURES             Location/Qualifiers
     source          1. .1312
                     /organism="Shigella flexneri"
                     /strain="5"
                     /db_xref="taxon:623"
BASE COUNT      326 a    328 c    338 g    320 t
ORIGIN
```

```
  Query Match             14.1%;  Score 216.8;  DB 1;  Length 1312;
  Best Local Similarity   52.5%;  Pred. No. 1.9e-48;
  Matches  526;  Conservative    0;  Mismatches  467;  Indels    9;  Gaps    2;
```

```
Qy    472 ATAACCAGCTGACCCGCCTGCCGCAAAGCCTCACGGGTCTGTCTTCAGCGGCACGCGTAT 531
Db     76 ATAATAATCTGACAACACTTCCGGAAAGTGTCCTGAGATTAGCTCAGAATGCCTTCGTAA 135

Qy    532 ATCTGGACGGGAATCCACTGTCTG-------TACGCACTCGTGACAGGCTCTGCGGACAT 584
Db    136 ATGTTGCAGGTAATCCATTGTCTGGCCATACCATGCGTACACTACAACAAATAACCACCG 195

Qy    585 CATTGGCCATTCAGGCATCAGGATACACTTCGATATGGCGGGGCCTTCCGTCCCCGGGAA 644
Db    196 GACCAGTTATTCTGGTCCTCGAATATTTTTCTCTATGGGAAATTCTGCC--ACAATTTCC 253

Qy    645 GCCCGGGCACTGCACCTGGCGGTCGCTGACTGGCTGACGTCTGCACGGGAGGGGGAAGCG 704
Db    254 GCTCCAGAACACTCCCTGGCTGATGCCGTGACAGCATGGTTCCCGGAAAACAAACAATCT 313

Qy    705 GCCCAGGCAGACAGATGGCAGGCGTTCGGACTGGAAGATAACGCCGCCGCCTTCAGCCTG 764
Db    314 GATGTATCACAGATATGGCATGCTTTTGAACATGAAGAGCATGCCAACACCTTTTCCGCG 373

Qy    765 GTCCTGGACAGACTGCGTGAGACGGAAAACTTCAAAAAAGACGCGGGCTTTAAGGCACAG 824
Db    374 TTCCTTGACCGCCTTTCCGATACCGTCTCTGCACGCAATACCTCCGGATTCCGTGAACAG 433

Qy    825 ATATCATCCTGGCTGACACAACTGGCTGAAGATGCTGCGCTGAGAGCAAAAACCTTTGCC 884
Db    434 GTCGCTGCATGGCTGGAAAAACTCAGTGCCTCTGCGGAGCTTCGACAGCAGTCTTTCGCT 493

Qy    885 ATGGCAACAGAGGCAACATCAACCTGCGAGGACCGGGTCACACATGCCCTGCACCAGATG 944
Db    494 GTTGCTGCTGATGCCACTGAGAGCTGTGAGGACCGTGTCGCGCTCACATGGAACAATCTC 553

Qy    945 AATAACGTACAACTGGTACATAATGCAGAAAAAGGGGAATACGACAACAATCTCCAGGGG 1004
Db    554 CGGAAAACCCTCCTGGTCCATCAGGCATCAGAAGGCCTTTTCGATAATGATACCGGCGCT 613

Qy   1005 CTGGTTTCCACGGGGCGTGAGATGTTCCGCCTGGCAACACTGGAACAGATTGCCCGGGAA 1064
Db    614 CTGCTCTCCCTGGGCAGGGAAATGTTCCGCCTCGAAATTCTGGAGGACATTGCCCGGGAT 673

Qy   1065 AAAGCCGGAACACTGGCTTTAGTCGATGACGTTGAGGTCTATCTGGCGTTCCAGAATAAG 1124
Db    674 AAAGTCAGAACTCTCCATTTTGTGGATGAGATAGAAGTCTACCTGGCCTTCCAGACCATG 733

Qy   1125 CTGAAGGAATCACTTGAGCTGACCAGCGTGACGTCAGAAATGCGTTTCTTTGACGTTTCC 1184
Db    734 CTCGCAGAGAAACTTCAGCTCTCCACTGCCGTGAAGGAAATGCGTTTCTATGGCGTGTCG 793

Qy   1185 GGCGTGACGGTTTCAGACCTTCAGGCTGCGGACGTTCAGGTGAAAACCGCTGAAAACAGC 1244
Db    794 GGAGTGACAGCAAATGACCTCCGCACTGCCGAAGCTATGGTCAGAAGCCGTGAAGAGAAT 853

Qy   1245 GGGTTCAGTAAATGGATACTGCAGTGGGGGCCGTTACACAGCGTGCTGGAACGCAAAGTG 1304
Db    854 GAATTTACGGACTGGTTCTCCCTCTGGGGACCATGGCATGCTGTACTGAAGCGTACGGAA 913

Qy   1305 CCGGAACGCTTTAACGCGCTTCGTGAAAAGCAAATATCGGATTATGAAGACACGTACCGG 1364
Db    914 GCTGACCGCTGGGCGCAGGCAGAAGAGCAGAAGTATGAGATGCTGGAGAATGAGTACTCT 973

Qy   1365 AAGCTGTATGACGAAGTGCTGAAATCGTCCGGGCTGGTCGACGATACCGATGCAGAACGT 1424
Db    974 CAGAGGGTGGCTGACCGGCTGAAAGCATCAGGTCTGAGCGGTGATGCGGATGCGCAGAGG 1033

Qy   1425 ACTATCGGAGTAAGTGCGATGGATAGTGCGAAAAAAGAATTT 1466
Db   1034 GAAGCCGGTGCACAGGTGATGCGTGAGACTGAACAGCAGATT 1075


RESULT   5
SHFIPAHY
LOCUS       SHFIPAHY     1375 bp    DNA             BCT       23-OCT-1991
DEFINITION  Shigella flexneri ipaH gene.
ACCESSION   M76444
NID         g152750
VERSION     M76444.1  GI:152750
KEYWORDS    LRG-like protein; ipaH gene.
SOURCE      Shigella flexneri (strain 5) DNA.
  ORGANISM  Shigella flexneri
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Shigella.
REFERENCE   1  (bases 1 to 1375)
  AUTHORS   Venkatesan,M.M., Buysse,J.M. and Hartman,A.B.
  TITLE     Sequence variation in two ipaH genes of Shigella flexneri 5 and
            homology to the LRG-like family of proteins
  JOURNAL   Mol. Microbiol. 5, 2435-2446 (1991)
  MEDLINE   92167809
FEATURES             Location/Qualifiers
     source          1. .1375
                     /organism="Shigella flexneri"
                     /strain="5"
                     /db_xref="taxon:623"
BASE COUNT      343 a    347 c    370 g    315 t
ORIGIN


  Query Match           14.1%;  Score 216.8;  DB 1;  Length 1375;
  Best Local Similarity   52.5%;  Pred. No. 1.9e-48;
  Matches  526;  Conservative    0;  Mismatches  467;  Indels    9;  Gaps    2;

Qy    472 ATAACCAGCTGACCCGCCTGCCGCAAAGCCTCACGGGTCTGTCTTCAGCGGCACGCGTAT 531
Db     76 ATAATAATCTGACAACACTTCCGGAAAGTGTCCTGAGATTAGCTCAGAATGCCTTCGTAA 135

Qy    532 ATCTGGACGGGAATCCACTGTCTG-------TACGCACTCGTGACAGGCTCTGCGGACAT 584
Db    136 ATGTTGCAGGTAATCCATTGTCTGGCCATACCATGCGTACACTACAACAAATAACCACCG 195

Qy    585 CATTGGCCATTCAGGCATCAGGATACACTTCGATATGGCGGGGCCTTCCGTCCCCGGGAA 644
Db    196 GACCAGTTATTCTGGTCCTCGAATATTTTTCTCTATGGGAAATTCTGCC--ACAATTTCC 253

Qy    645 GCCCGGGCACTGCACCTGGCGGTCGCTGACTGGCTGACGTCTGCACGGGAGGGGGAAGCG 704
Db    254 GCTCCAGAACACTCCCTGGCTGATGCCGTGACAGCATGGTTCCCGGAAAACAAACAATCT 313

Qy    705 GCCCAGGCAGACAGATGGCAGGCGTTCGGACTGGAAGATAACGCCGCCGCCTTCAGCCTG 764
Db    314 GATGTATCACAGATATGGCATGCTTTTGAACATGAAGAGCATGCCAACACCTTTTCCGCG 373

Qy    765 GTCCTGGACAGACTGCGTGAGACGGAAAACTTCAAAAAAGACGCGGGCTTTAAGGCACAG 824
Db    374 TTCCTTGACCGCCTTTCCGATACCGTCTCTGCACGCAATACCTCCGGATTCCGTGAACAG 433

Qy    825 ATATCATCCTGGCTGACACAACTGGCTGAAGATGCTGCGCTGAGAGCAAAAACCTTTGCC 884
Db    434 GTCGCTGCATGGCTGGAAAAACTCAGTGCCTCTGCGGAGCTTCGACAGCAGTCTTTCGCT 493

Qy    885 ATGGCAACAGAGGCAACATCAACCTGCGAGGACCGGGTCACACATGCCCTGCACCAGATG 944
Db    494 GTTGCTGCTGATGCCACTGAGAGCTGTGAGGACCGTGTCGCGCTCACATGGAACAATCTC 553

Qy    945 AATAACGTACAACTGGTACATAATGCAGAAAAAGGGGAATACGACAACAATCTCCAGGGG 1004
Db    554 CGGAAAACCCTCCTGGTCCATCAGGCATCAGAAGGCCTTTTCGATAATGATACCGGCGCT 613

Qy   1005 CTGGTTTCCACGGGGCGTGAGATGTTCCGCCTGGCAACACTGGAACAGATTGCCCGGGAA 1064
Db    614 CTGCTCTCCCTGGGCAGGGAAATGTTCCGCCTCGAAATTCTGGAGGACATTGCCCGGGAT 673

Qy   1065 AAAGCCGGAACACTGGCTTTAGTCGATGACGTTGAGGTCTATCTGGCGTTCCAGAATAAG 1124
Db    674 AAAGTCAGAACTCTCCATTTTGTGGATGAGATAGAAGTCTACCTGGCCTTCCAGACCATG 733

Qy   1125 CTGAAGGAATCACTTGAGCTGACCAGCGTGACGTCAGAAATGCGTTTCTTTGACGTTTCC 1184
```

```
Db    734 CTCGCAGAGAAACTTCAGCTCTCCACTGCCGTGAAGGAAATGCGTTTCTATGGCGTGTCG 793

Qy   1185 GGCGTGACGGTTTCAGACCTTCAGGCTGCGGACGTTCAGGTGAAAACCGCTGAAAACAGC 1244

Db    794 GGAGTGACAGCAAATGACCTCCGCACTGCCGAAGCTATGGTCAGAAGCCGTGAAGAGAAT 853

Qy   1245 GGGTTCAGTAAATGGATACTGCAGTGGGGGCCGTTACACAGCGTGCTGGAACGCAAAGTG 1304

Db    854 GAATTTACGGACTGGTTCTCCCTCTGGGGACCATGGCATGCTGTACTGAAGCGTACGGAA 913

Qy   1305 CCGGAACGCTTTAACGCGCTTCGTGAAAAGCAAATATCGGATTATGAAGACACGTACCGG 1364

Db    914 GCTGACCGCTGGGCGCAGGCAGAAGAGCAGAAGTATGAGATGCTGGAGAATGAGTACTCT 973

Qy   1365 AAGCTGTATGACGAAGTGCTGAAATCGTCCGGGCTGGTCGACGATACCGATGCAGAACGT 1424

Db    974 CAGAGGGTGGCTGACCGGCTGAAAGCATCAGGTCTGAGCGGTGATGCGGATGCGCAGAGG 1033

Qy   1425 ACTATCGGAGTAAGTGCGATGGATAGTGCGAAAAAAGAATTT 1466

Db   1034 GAAGCCGGTGCACAGGTGATGCGTGAGACTGAACAGCAGATT 1075


RESULT   6
AE000074
LOCUS       AE000074    14263 bp    DNA             BCT       12-DEC-1997
DEFINITION  Rhizobium sp. NGR234 plasmid pNGR234a, section 11 of 46 of the
            complete plasmid sequence.
ACCESSION   AE000074 U00090
NID         g2182401
VERSION     AE000074.1  GI:2182401
KEYWORDS    .
SOURCE      Rhizobium sp. NGR234.
  ORGANISM  Rhizobium sp. NGR234
            Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
            Rhizobiaceae; Rhizobium.
REFERENCE   1  (bases 1 to 14263)
  AUTHORS   Freiberg,C., Fellay,R., Bairoch,A., Broughton,W.J., Rosenthal,A.
            and Perret,X.
  TITLE     Molecular basis of symbiosis between Rhizobium and legumes
  JOURNAL   Nature 387 (6631), 394-401 (1997)
  MEDLINE   97305956
REFERENCE   2  (bases 1 to 14263)
  AUTHORS   Freiberg,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-NOV-1996) Genome Analysis, Institute for Molecular
            Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
REFERENCE   3  (bases 1 to 14263)
  AUTHORS   Freiberg,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-DEC-1997) Genome Analysis, Institute for Molecular
            Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
  REMARK    Update by submitter
FEATURES             Location/Qualifiers
     source          1. .14263
                     /organism="Rhizobium sp. NGR234"
                     /plasmid="pNGR234a"
                     /strain="NGR234"
                     /db_xref="taxon:394"
     source          <1. .6634
                     /organism="Rhizobium sp. NGR234"
                     /db_xref="taxon:394"
                     /clone="pXBS23"
     source          <1. .>14263
                     /organism="Rhizobium sp. NGR234"
                     /db_xref="taxon:394"
                     /clone="pXBS4"
     misc_feature    131. .235
                     /note="mosaic sequence homolog (MSH) 14a: similar to
                     RIME1"
                     /note="Region: MSH14a"
     gene            548. .2188
                     /gene="y4fR"
     CDS             548. .2188
                     /gene="y4fR"
                     /note="hypothetical 60.5 Kd protein, homologous to
                     Shigella flexneri 60 Kd antigen (IpaH)"
                     /codon_start=1
                     /transl_table=11
                     /product="Y4fR"
                     /protein_id="AAB91674.1"
                     /db_xref="PID:g2182402"
                     /db_xref="GI:2182402"
                     /translation="MNVQRPGLAVGPLFENPESESSEPGSPAAAARWVEASTEAEASA
                     ASSSQGQIVAAPTAEERPWEGRPQEAVSRTRAWREAGDVDEPLDLSFLSLTPLSIPLV
                     SGLRRLNVNNNQLGDLPDTLPGTLLELEASENRLTRLPDLPAGLQRLNVENNRLTNLP
                     EPLPAALEWLGAGYNQLTRLPEMIPPELIWLGARNNQLTSVPESLLTQLGQWSSIDLE
                     NNPLPHGVQTNLVTAMHAAGYAGPQIFLPMGPVELARRPLHEVVADWLEGDLETVAAW
                     RGFANEQGARDYAHFLDRLRTTVNYGNDAFRQAVAIGLRQAVARPQLRAQYFEQASGA
                     SDSCEDRITLTWNGMQTALLIADVEDGVYDGSLHQLLQHGRVMFRLEALDGIARETVN
                     SLRRTDPDADIDEIEVYLAYQTQLRDTLELRHVAPDMRFLNVSHVTEEDVARAASSVR
                     ELEARGFGEYVATRWQPWERVMRRIAPASHAAMQEQLIEAMGEEFRSRLDEKLAEHGL
                     TGDADAERVFGAEILNDIARRIKGETMEKVLRGRGLEL"
     gene            complement(2541. .3104)
                     /gene="y4gA"
     CDS             complement(2541. .3104)
                     /gene="y4gA"
                     /note="hypothetical 20.9 Kd protein; low similarity to
                     Y4rE"
                     /codon_start=1
                     /transl_table=11
                     /product="Y4gA"
                     /protein_id="AAB91675.1"
                     /db_xref="PID:g2182403"
                     /db_xref="GI:2182403"
                     /translation="MCQWAATKGTRNHAYEREKLERAFAKILRHRDPSVPPRRSDPRP
                     RKEVARQWRKPHIYSPADLRRMLDIARSYPSPRAALRQENMYTMLLLAYCAGLRRGEL
                     ARLDLGDVNLGDGTITVRQTKFFKTRILPLPDSVVVSFGPISLHGVGSAHSRLLTPPC
                     SGTNKAVPAPRQEGSPGCLQTSYVVPG"
     misc_feature    2541. .3104
                     /note="insertion sequence homolog (ISH) 12f: region
                     including a putative IS element-derived ORF (y4gA)"
                     /note="Region: ISH12f"
     gene            complement(3220. .3639)
                     /gene="y4gB"
     CDS             complement(3220. .3639)
                     /gene="y4gB"
                     /note="hypothetical 16.1 Kd protein"
                     /codon_start=1
                     /transl_table=11
                     /product="Y4gB"
                     /protein_id="AAB91676.1"
                     /db_xref="PID:g2182404"
                     /db_xref="GI:2182404"
                     /translation="MTARWPDPDRAYIGRYVASLDLRSIKSRTCYRQVLHGFQDVVER
                     HEALDQQVLLAWLRQSSDRWAATTLLHRTRIIDRFLDHFLEAAAIDHNPVEDLREACH
                     IKQCMPVWRALISCDPEQALAQLRQPEPFGSAGRDHG"
     misc_feature    complement(3635. .4215)
                     /note="insertion sequence homolog (ISH) 12e: region
                     including an integrase/recombinase (y4gC)"
                     /note="Region: ISH12e"
     gene            complement(3636. .4214)
                     /gene="y4gC"
     CDS             complement(3636. .4214)
                     /gene="y4gC"
                     /note="putative integrase/recombinase"
                     /codon_start=1
                     /transl_table=11
                     /product="Y4gC"
                     /protein_id="AAB91677.1"
                     /db_xref="PID:g2182405"
                     /db_xref="GI:2182405"
                     /translation="MPSILERDQIAAVLVSASRDKTPAGLRDHAILQLLATYGLRSGE
                     IRNMRIEDIDWRTETIRVRHSKTQAYTLLPLTEPVGEAILVYLRSGRPATDARELFVR
                     TRAPYRKLDKLYSMVRRRLRDAGVKPRGKRGPHIFRHARATELLRAAVPQKIIGELLG
                     HRSIASTAPYLKLATEDLRAIALDLPGTEVSA"
```

```
source          3683. .>14263
                /organism="Rhizobium sp. NGR234"
                /db_xref="taxon:394"
                /clone="pXB285"
gene            complement(4314. .4592)
                /gene="y4gD"
CDS             complement(4314. .4592)
                /gene="y4gD"
                /note="hypothetical 10.5 Kd protein"
                /codon_start=1
                /transl_table=11
                /product="Y4gD"
                /protein_id="AAB91678.1"
                /db_xref="PID:g2182406"
                /db_xref="GI:2182406"
                /translation="MQDRWPPAPKVTCAADALRFAICNEYETWLREERGLARASIDAF
                LWEARHFLGWQLERCGAEGLMEVSVGDIDRYVDMLASKLTRSSLKCAA"
misc_feature    5350. .5398
                /note="nod box-like sequence"
                /note="Region: nod box 5"
                /evidence=not_experimental
misc_feature    6735. .7535
                /note="insertion sequence homolog (ISH) 15"
                /note="Region: ISH15"
gene            6756. .7502
                /gene="y4gE"
CDS             6756. .7502
                /gene="y4gE"
                /note="hypothetical 27.7 Kd protein (possibly
                fragmentous)"
                /codon_start=1
                /transl_table=11
                /product="Y4gE"
                /protein_id="AAB91679.1"
                /db_xref="PID:g2182407"
                /db_xref="GI:2182407"
                /translation="MRAGQGGSLRLSVCSVIVQAPSRRALDAKARHAVRDDTILRQLK
                RDAAPRSLIATPQARQVADRFHLYADLRVAIEEQMSLSGRARGRALLPDKIIGNAQVD
                LIQDDPHVDATHRRRVRHGHRQSRQVVFETVHALRKKGLSCSAIARRTGYGRRSIAKW
                LIFETPPDRRKSVLKPTSPIKDGNRCGRHLYTISNSAVTSAVSRISSDFSQPGAAPRG
                RARTVRRRLLSYPINRLAMLSRCGIRRLAT"
gene            8469. .9521
                /gene="y4gF"
CDS             8469. .9521
                /gene="y4gF"
                /note="probable dTDP-D-glucose-4,6-dehydratase"
                /codon_start=1
                /transl_table=11
                /product="Y4gF"
                /protein_id="AAB91680.1"
                /db_xref="PID:g2182408"
                /db_xref="GI:2182408"
                /translation="MRILVTGGAGFIGSALVRYLVSINAEVLNVDKLTYAGNLASLKP
                VEGLRNYRFLRADICDRVAINEAFETFQPDYVIHLAAESHVDRSITGADDFVQTNVNG
                TFTMLETARQYWSNLSQNRKAFFKMLHVSTDEVYGSLGDRGQFEEVSPYDPSSPYSAS
                KAASDHFATAWQRTYGLPVVISNCSNNYGPFHFPEKLIPLMILNALDRKPLPVYGTGS
                NIRDWLYVDDHARALWLIVREGRPGEKYNVGGRNELRNIDVVNRICLLLDELSPNASH
                YGDLITFVKDRPGHDARYAIDATKLETELGWKAQENFDTGIRKTVEWYLENGWWWQPL
                RDKVYSGERLGLLEKA"
gene            9525. .10415
                /gene="y4gG"
CDS             9525. .10415
                /gene="y4gG"
                /note="probable dTDP-4-dehydrorhamnose reductase"
                /codon_start=1
                /transl_table=11
                /product="Y4gG"
                /protein_id="AAB91681.1"
                /db_xref="PID:g2182409"
                /db_xref="GI:2182409"
                /translation="MRLAVTGKNGQIALALKAQARPDVEILTLGRPNFDLACRSTVAS
                SIRDAAPDIIVSLAAYTAVDKAESEPYEAFAVNRDGVQALAEAAAGLGVPVIHLSTDY
                VFDGAKPVPYCEEDRTGPISVYGRSKLEGEFAVASANPNHTILRTSWVYSRYGQNFVK
                KMLRLADTNDELNVVADQLGCPTSADDISVAVMTIARRMLSSSSADLRGIFHLSGSGE
                ASWAAFAKYVFSVYDEITGRQIKVHDISAAEYPTPARRPANSRLHCDKLERTFGIRLP
                NWEESTRRLVWALLLEGKDA"
gene            10412. .11272
                /gene="y4gH"
CDS             10412. .11272
```

Query Match 9.3%; Score 142.6; DB 2; Length 14263;
Best Local Similarity 50.4%; Pred. No. 2.8e-28;
Matches 609; Conservative 0; Mismatches 559; Indels 41; Gaps 9;

```
Qy    256 ACGTTGCCGCCAGGACTAGGGGATCTGGCGGTCTCTAATAACCAGCTGACCAGCCTGCCG 315
Db    908 ACCCTCCCGGGAACGCTTCTGGAGCTGGAGGCCAGCGAGAACCGGCTGACCCGCCTGCCC 967

Qy    316 GAGATGCCGCCAGCACTACGGGAGCTGAGGGTCTCTGGTAACAACCTGACCAGCTGCGCG 375
Db    968 GACCTTCCTGCCGGGCTCCAGAGGCTGAACGTCGAAAACAACCGGCTGACCAATCTCCCC 1027

Qy    376 ----CGCTGCCGTCAGGACTACAGAAGCTGTGGGCCTATAATAATCGGCTGACCAGCCTG 431
Db   1028 GAGCCCCTCCCGGCCGCGCTCGAGTGGCTGGGTGCTGGCTACAACCAGCTAACCCGCCTG 1087

Qy    432 CCGGAGATG---TCGCCAGGACTACAGGAGCTGGATGTCTCTCATAACCAGCTGACCCGC 488
Db   1088 CCCGAGATGATTCCCCCCGAGCTTATATGGCTGGGCGCCAGAAACAATCAGCTGACCAGC 1147

Qy    489 CTGCCGCAAAGCCTCACGGGTCTGTCTTCAGCGGCACGCGTATATCTGGACGGGAATCCA 548
Db   1148 GTGCCCGAAAGCCTGCTGACGCAG--CTAGGCCAATGGTCTAGCATTGATTTGGAAAATA 1205

Qy    549 CTGTCTGTACGCACTCGTGACAGGCTCTGCGGACATCATTGGCCATTCAGGCATCAGGAT 608
Db   1206 ATCCGCTGCCGCATGGGGTACAAACAAACTTGGTGACA---GCCATGCATGCCGCGGGTT 1262

Qy    609 ACACTTCGATATGGCGGGGCCTTCCGTCCCCGGGAAGCCCGGGCACTGCAC--CTGGCGG 666
Db   1263 ACGCA---GGACCGCAGATCTTTTTGCCTATGGGACCGGTGGAACTTGCACGACGGCCCC 1319

Qy    667 TCGCTGACTGGCTGACGTCTGCACGGGAGGGGGAAGCGGCCCAGGCAGACAGATGGCAGG 726
Db   1320 TTCACGAGGTTGTCGCGGATTGGCTCGAGGGGGATTTGGAAACGGTGGCCGCCTGGCGGG 1379

Qy    727 CGTTCGGACTGGAAGATAACGCCGCCGCCTTCAGCCTGGTCCTGGACAGACTGCGTGAGA 786
Db   1380 GCTTCGCCAACGAACAAGGTGCCCGGGATTATGCACATTTCCTCGACCGGCTCCGTACCA 1439

Qy    787 CGGAAAACTTCAAAAAAGACGCGGGCTTTAAGGCACAGATATCATCCTGGCTGACACAAC 846
Db   1440 CCGTGAACTATGGCAATGATGCGTTCCGGCAGGC---GGTGGCCATAGGTCTGCGGCAGG 1496

Qy    847 TGGCTGAAGATGCTGCGCTGAGAGCAAAAACCTTTGCCATGGCAACAGAGGCAACATCAA 906
Db   1497 CAGTGGCCAGGCCGCAATTGCGCGCACAGTATTTTGAGCAGGCTTCCGGAGCTAGCGATA 1556

Qy    907 CCTGCGAGGACCGGGTCACACATGCCCTGCACCAGATGAATAACGTACAACTGGTACATA 966
Db   1557 GCTGTGAGGATCGCATTACTTTGACCTGGAATGGAATGCAGACCGCACTTCTGATCGCTG 1616

Qy    967 ATGCAGAAAAAGGGGAATACGACAACAATCTCCAGGGGCTGGTTTCCACGGGGCGTGAGA 1026
Db   1617 ATGTAGAGGACGGGGTCTATGACGGGTCGCTCCACCAGCTTCTCCAGCACGGCCGTGTCA 1676

Qy   1027 TGTTCCGCCTGGCAACACTGGAACAGATTGCCCGGGAAAAAGCCGGAACACTG------- 1080
Db   1677 TGTTCCGTTTGGAGGCACTGGACGGGATCGCGCGCGAGACGGTCAACTCGCTTCGACGTA 1736

Qy   1080 ----------GCTTTAGTCGATGACGTTGAGGTCTATCTGGCGTTCCAGAATAAGCTGA 1128
Db   1737 CCGATCCGGACGCCGACATCGACGAGATCGAGGTCTATCTTGCCTACCAGACACAGCTGC 1796

Qy   1129 AGGAATCACTTGAGCTGACCAGCGTGACGTCAGAAATGCGTTTCTTTGACGTTTCCGGCG 1188
```

```
Db   1797 GCGACACCCTGGAGCTGCGGCACGTCGCCCCAGACATGCGCTTCTTGAACGTCTCTCACG 1856

Qy   1189 TGACGGTTTCAGACCTTCAGGCTGCGGACGTTCAGGTGAAAACCGCTGAAAACAGCGGGT 1248

Db   1857 TTACCGAGGAGGACGTTGCCAGAGCTGCCAGTTCCGTGCGGGAACTTGAGGCGAGGGGAT 1916

Qy   1249 TCAGTAAAT---GGATACTGCAGTGGGGGCCGTTACACAGCGTGCTGGAACGCAAAGTGC 1305

Db   1917 TCGGCGAGTATGTGGCAACGCGCTGGCAACCTTGGGAAAGAGTCATGAGGCGCATCGCCC 1976

Qy   1306 CGGAACGCTTTAACGCGCTTCGTGAAAAGCAAATATCGGATTATGAAGACACGTACCGGA 1365

Db   1977 CCGCAAGCCATGCAGCGATGCAGGAGCAGCTCATCGAAGCGATGGGAGAGGAGTTCAGAA 2036

Qy   1366 AGCTGTATGACGAAGTGCTGAAATCGTCCGGGCTGGTCGACGATACCGATGCAGAACGTA 1425

Db   2037 GCCGCTTGGACGAAAAGCTCGCCGAGCACGGCCTGACGGGTGATGCCGATGCCGAACGGG 2096

Qy   1426 CTATCGGAG 1434

Db   2097 TATTCGGAG 2105
```

```
RESULT   7
I66494/c
LOCUS       I66494       7218 bp    DNA             PAT       23-DEC-1997
DEFINITION  Sequence 14 from patent US 5670367.
ACCESSION   I66494
NID         g2724471
VERSION     I66494.1  GI:2724471
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 7218)
  AUTHORS   Dorner,F., Scheiflinger,F. and Falkner,F.Gunter.
  TITLE     Recombinant fowlpox virus
  JOURNAL   Patent: US 5670367-A 14 23-SEP-1997;
FEATURES             Location/Qualifiers
     source          1. .7218
                     /organism="unknown"
BASE COUNT     1944 a   1491 c   1486 g   1929 t    368 others
ORIGIN
```

```
  Query Match             3.0%;  Score 45.4;  DB 5;  Length 7218;
  Best Local Similarity   8.4%;  Pred. No. 0.055;
  Matches   37;  Conservative  210;  Mismatches  196;  Indels    0;  Gaps    0;

Qy    552 TCTGTACGCACTCGTGACAGGCTCTGCGGACATCATTGGCCATTCAGGCATCAGGATACA 611

Db   1520 TCTATTCAGTTTCAAAAAACGGCATGTAGGCATCACTGTAATTACCTATCTATGCAAGTA 1461

Qy    612 CTTCGATATGGCGGGGCCTTCCGTCCCCGGGAAGCCCGGGCACTGCACCTGGCGGTCGCT 671

Db   1460 GTTAAAGAGATAGAAGAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1401

Qy    672 GACTGGCTGACGTCTGCACGGGAGGGGGAAGCGGCCCAGGCAGACAGATGGCAGGCGTTC 731

Db   1400 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1341

Qy    732 GGACTGGAAGATAACGCCGCCGCCTTCAGCCTGGTCCTGGACAGACTGCGTGAGACGGAA 791

Db   1340 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1281

Qy    792 AACTTCAAAAAAGACGCGGGCTTTAAGGCACAGATATCATCCTGGCTGACACAACTGGCT 851

Db   1280 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1221

Qy    852 GAAGATGCTGCGCTGAGAGCAAAAACCTTTGCCATGGCAACAGAGGCAACATCAACCTGC 911

Db   1220 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1161
```

```
Qy    912 GAGGACCGGGTCACACATGCCCTGCACCAGATGAATAACGTACAACTGGTACATAATGCA 971

Db   1160 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1101

Qy    972 GAAAAAGGGGAATACGACAACAA 994

Db   1100 RRRRRRRRRRRRRRRRRRRRRR 1078
```

```
RESULT   8
AF027770
LOCUS       AF027770    30683 bp    DNA             BCT       03-DEC-1998
DEFINITION  Mycobacterium smegmatis FxbA (fxbA) gene, partial cds; FxbB (fxbB),
            FxbC (fxbC), and FxuD (fxtD) genes, complete cds; and unknown
            genes.
ACCESSION   AF027770
NID         g3560502
VERSION     AF027770.1  GI:3560502
KEYWORDS    .
SOURCE      Mycobacterium smegmatis.
  ORGANISM  Mycobacterium smegmatis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium.
REFERENCE   1  (bases 1 to 30683)
  AUTHORS   Fiss,E.H., Yu,S. and Jacobs,W.R. Jr.
  TITLE     Identification of genes involved in the sequestration of iron in
            mycobacteria: the ferric exochelin biosynthetic and uptake pathways
  JOURNAL   Mol. Microbiol. 14 (3), 557-569 (1994)
  MEDLINE   95191405
REFERENCE   2  (bases 1 to 30683)
  AUTHORS   Yu,S., Fiss,E. and Jacobs,W.R. Jr.
  TITLE     Analysis of the exochelin locus in mycobacterium smegmatis:
            biosynthesis genes have homology with genes of the peptide
            synthetase family
  JOURNAL   J. Bacteriol. 180 (17), 4676-4685 (1998)
  MEDLINE   98389687
REFERENCE   3  (bases 1 to 30683)
  AUTHORS   Yu,S. and Jacobs,W.R. Jr.
  TITLE     Genetic analysis of iron uptake in Mycobacterium smegmatis
  JOURNAL   Unpublished
REFERENCE   4  (bases 1 to 30683)
  AUTHORS   Yu,S. and Jacobs,W.R. Jr.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-SEP-1997) Microbiology & Immunology, Albert Einstein
            College of Medicine, 1300 Morris Park Ave, Bronx, NY 10461, USA
FEATURES             Location/Qualifiers
     source          1. .30683
                     /organism="Mycobacterium smegmatis"
                     /strain="mc2155"
                     /db_xref="taxon:1772"
     gene            complement(<1. .429)
                     /gene="fxbA"
     CDS             complement(<1. .429)
                     /gene="fxbA"
                     /note="similar to methionyl-tRNA formyltransferase"
                     /codon_start=1
                     /transl_table=11
                     /product="FxbA"
                     /protein_id="AAC82546.1"
                     /db_xref="PID:g3560503"
                     /db_xref="GI:3560503"
                     /translation="MWGSLKPPPNLPESWVFPVRVVMFGYQTWGHRTLQALLKSRHEV
                     CLVVTHPTSDHAYESIWADSVEDLARGAGIEVLLAKRPTPELVERVRELAPDVGVANN
                     WRTRLPRELFSIPKYGTVNLHDSLLPKFTGFSPVIWSLIER"
     CDS             811. .2535
                     /note="similar to ABC transporters; Orf1"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAC82547.1"
                     /db_xref="PID:g3560504"
                     /db_xref="GI:3560504"
```

```
                     /translation="MADTTRTDLSWTGVRVLRRTVSRNLQVAHVRYDAHRPAPAVRGV
                     GTGSDRHHRRPRGRDRQRRTILRWIAVLAALFVVLTVVYRFGARLLMFAIARESHLLR
                     VESSAKILDPLGIETDYKVGELLSISSDDADEVSYLLDYVPRIVGAVVGTVVCGAVLL
                     SIHLPLGLMVLIGVPVVVFGLQLTAPMIARRVEDQQAEIGRATAWPPTSSADSAVAGI
                     GAQANAAARYRVASRRALTATLRAARIQSIHSGAAAAAGALAAMAVAVVAAYFAIRGS
                     MIVGQLITVIGLAQFLIEPFSLLAVVPSWVAEARASANRVARVLSAPTRHSPVFAADP
                     RDAAARLSVRNATHAGLKNLSFDVEPGEFVAVLTSDVRDAGALIDLLSGYQRAEDKGT
                     VLVGGRRLDQIPLKERREHLLVEHHLSSLFSGTLETNLNVTGTDPRRGRVSVDDALQA
                     SCAVDVVDLHPDGLAHAVVERGASLSGGQRQRWTLARALLADPAVLVLHDPTTAIDTV
                     TEQVIADGIRRLRTRAGRTTVVLTSSPALLAAADRVLLVIDGRLDSEGTHRDLVDAHS
                     DYRDLVTR"
CDS                  2532. .4301
                     /note="similar to ABC transporters; Orf2"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAC82548.1"
                     /db_xref="PID:g3560505"
                     /db_xref="GI:3560505"
                     /translation="MISDEPVTPPVLPVATARRSGAWLIADLRGRRTALAAVVTVGLA
                     AAAASVVPIYLLGMLVDRVEHGDGTRGLVALGTVIALAAVAGGLGTGLSTYLTSRLGE
                     QMLADLRERVLERALNLPATLIEESGRGDLLSRVGPDVAVVARTVAQVLPVILNGFFL
                     GVVTLVGMTTLDWRLGLAGALAIPAYIAGLHWYLPRSAPMYADERIAIANRAQVTVES
                     IQGARTIDAYEIHDRHLADIDRASGRARDISIAVFTFFTRLVGRVNRAEFIGLAATLV
                     VGFLLVRSGETTVGQVTAAALMFHRLFNPVGELMYNFDEIQSASASLARLVGVIDLDA
                     PGPDARRTAPARPHGADISVENVSFGYGTGPEVVHGVSLHVPAGTRCALVGTTGAGKT
                     TLAAIVAGLLTPTSGHARIGGVPVTEIADLRRWVATITQEVHVFSGPLIDDLRLVAPQ
                     ASVDDVWTALETVGASTWVRALDDGLDTVVGEHGHELTAAQAQHVAMARLVLADPEVV
                     VLDEATAEAGSAHAAELEDAASAATAGRTTLIVAHRLTQAAQADQVAVMADGRIVENG
                     THAELIALGGRYARLWQAWIRHN"
gene                 4335. .12044
                     /gene="fxbB"
CDS                  4335. .12044
                     /gene="fxbB"
                     /note="similar to peptide synthetases"
                     /codon_start=1
                     /transl_table=11
                     /product="FxbB"
                     /protein_id="AAC82549.1"
                     /db_xref="PID:g3560506"
                     /db_xref="GI:3560506"
                     /translation="MTADSLDIAELLELWNNHSTPERTSTVPTLFAAQCALTPDEVAV
                     VDGERRLTYRHLETHVAQLAHAVRVAAGEGPEPIVAIGVPRSAEMVVCVLAAMMAGVA
                     FVPLDPSWPAHGRRQVLADSGAVATFITREDESDWGVPGLRVDLGAWQFTAESPVLPQ
                     ADVHPAQLAYVIFTSGSTGKPKGAMIRQDAIAERLQWQRDHILHFGKHDHTDASLFKA
                     PLSFDISVNEILLPLVSGGRVVVAVPDGEKDPEYLLELIRTEQVTFVYLVSSMLDTLL
                     ELDRLATADGAPSSLASLRHVWCGGEVLTPGLFARFRKQLTTTLYHGYGPAEATIGVS
                     HVIYRDTAERIATSIGRPNPHTQLYVLDEYLRPVPPGVGGELYAAGFLLGRGYVNAPS
                     LTASRFVANPFDGNGSRMYRTGDLARWTEDGSLEFLGRADNQVKIGGRRVELEEIESQ
                     LADHPAVRHAVVDVHRQGGADVLVGYLVAADGVRNDAAWHAEVADWARTRLPEYMVPK
                     AFVALDQVPLTANGKTDRRALPAPDLERSGTVKPPRTPRETVLCQVFADEIDAVGVDE
                     DFFALGGDSIVAIRVVSRLRAAGYTLRPRDMFAHRTVEALAPLLGDSDVRDTGPAVDP
                     TGAATPTPILRWLDEVGTAGSVLNGFHQGMSLVTPADADENTLRAAIAATVRRHHVLW
                     APPGRTASDIDIPGTPPETRLLIADASDGIPAQAEKVARQLVSLLDTARIAFGWIRRP
                     AAPGRLVVIADHTVIDGVSLRILAEDIATAYGLIAEGRAVELPTPHTSWRAWAQRLAD
                     TAAAGGFDADLEHWQQVCATTETPWGDRALDPAIDTVATESRLTVELPSAVTDAILTT
                     VPDRIHGHVNDALVAALYLALRRWLHTRGVGADTLLVEMEGHGREGHLVDSDTAGLDL
                     SNTVGWFTTLYPVALRDAEFDWQAAVSGGPQLGAAVRSVKDQLRSVPSHGFSYGALRY
                     LRDGTSGLEAAPQVLFNYLGRFGTADRPWALANDTTAVLEDRDPGMPLPRLLEVNAEA
                     VTTADGSVLRATFSWPAHAVAEVDVRTLAGMWTDLLTAIASSDDVRGHSASDFDRVSV
                     TADDVAELERRYPGLTDLLPLTPTQQGIYFHSTFSRDRDPYVVQQIVDITGPLDTERF
                     QRATESVAARHRALGAVFTTLSDGTPVAVHAATVAPDFDVLDARHAADPSTVVAQRAR
                     WERERRFDLAAAPLTRYTLVRRRDDLHTMIETVHHIVADGWSVPIVLDDLLTAYAGDD
                     FDGPAPEFARFVDWLEQRDTDSDRAAWAPALEDITEPTRLAAADGARGRTTGSGFGTR
                     TVTLQSRSAVADAANGAAVTVGTLLHTAWGLALGRLTGRDDVVFGTVVSGRGADVHGI
                     ESMVGLLVNTVPVRVAWSADDTGADVAGRLAAVESTLLEHHHLPLTEAHRLAGVGELF
                     DTLVVIENLGATTHTRGDLTLGDIGVIEAPHYPLTLMISVRDTITVTVTNDREHVSDV
                     LADTAVAAFTEVLTALTADPGVSPGDIALPAPQATAPPTQAPQTVTDLIAAAIAEHRD
                     DIALVVGETEWTYGQLGARAGELAAALAEAGVRRGDIVALATARSADLVAAIWAIIAA
                     GAAYLRSTRLSADRIEYMLRHARPTAVIADGVGAHVVSGALPADTIVVSTTATHAAVP
                     FTPVPVDGADAVSVLYTSGSTGEPKAVVGTHAALANRLAWAVEAWPAATRIAKSPLSF
                     IDGTTELLAGLAAGARTVLAGDETARDGRRLAQLVAAHGVEQLLAVPSLAAVLADERT
                     EDVAELNRWIVSGEALEPRHLHALRTACPTAEIVNSYGSSEVAGDVLAGVQDDAGITL
                     GAAVPGAGIRILDSRLRQLPAGVIGEIYVTGGQLARGYLGRPGQTATRFVAAPGGERM
                     YRTGDLGALLPGGRVVFAGRADDQLKINGHRVEPGEIESVLARQPGVREAAVIGTERI
                     RAFVVLESDAPGAGDLLTAVSAELPGHLVPSSLRPVDAIPLLPNGKRDNNALRSLLVP
                     GESTGTAPTNDVERAVLDVMTGVLGRDPLGADDDFFAVGGDSISAIRVTSRLARAGYH
                     IATEDVFRGRTAAGVAALVDTVESLAHDAAIEPFATVRLSPQTIDRIRESGQVEDIWA
                     MSPLQLGVYYQSTLADAGPTYIAQNVFEFDRRIDVDAMRRAFTALLRRHPQLRAGFRT
                     VEHAEPEPAPDATPVVQVVVTDPPSDLTVVDLTDQDDPAAAQHIVDTDRTAPFDVATP
                     PLLRLTVIRLPGGRDRMLFTYHFLLFDGWSRELVLRDLFALYDSDAQHGAIAPHGDLV
                     VRHLQWLATDGDPQASGARDAWRDLLAGLTEPTLASGVSPDHPDARPGTEPGRIVVRV
                     PEDVTTQLHTCATAHGATLNSVVTAAVALVTGYHAGTIDVVIGTTVAGRPGHLVGIDE
                     TIGLFLNTVPVRVDMSPTRSVAAAMASIGEQRVAMMRHDHLGLGQIQRAAGDSGSALF
                     DSLLVLQNFLDDDTFTDLESRHGYRSTLSYHDTTHFPLTWVLTPGRRNSRSSSSTAWS
                     TTPGPPRWSNSSSRCCGPSPRNPIPPWARST"
gene                 11969. .26899
                     /gene="fxbC"
CDS                  11969. .26899
                     /gene="fxbC"
                     /note="similar to peptide synthetases"
                     /codon_start=1
                     /transl_table=11
                     /product="FxbC"
                     /protein_id="AAC82550.1"
                     /db_xref="PID:g3560507"
                     /db_xref="GI:3560507"
                     /translation="MVEQLVTVLRTIAAEPDTAVGAVDLIGADRRAQLGAQVVVHRTT
                     ARAGHDRRTPRAACRTESRRRRAGVRRAVGDLPRIRRPRLAIRAPSARTRRGARDVRR
                     AGTASIDRHGGRAVRGAARRCGLSAARAGPPHRPPPHDHRQRRTRAARHDDRPHGTDR
                     ARTSPRCRCHRARRRGNRCHTRRYARAPADRRRTRGFRVRQHAPEHPAYLIYTSGSTG
                     RPKGVLTGYAGLTNMYFNHREAIFAPTVARAGSAEQLRIAHTVSFSFDMSWEELFWLV
                     EGHQVHVCDEELRRDAPALVAYCHRHRIDVINVTPTYAHHLFDAGLLDDGAHTPPLVL
                     LGGEAVGDGVWSALRDHPDSAGYNLYGPTEYTINTLGGGTDDSDTPTVGQPIWNTRGY
                     ILDAALRPVPDGAVGELYIAGTGLALGYHRRAGLTAATMVADPYVPGGRMYRTGDLVR
                     RRPGSAELLDYLGRVDDQVKIRGYRVELGEIESVLTRADGVARCAVVARATGANPPVK
                     TLAAYVIPDRWPAEDAAFITGLRDHLARVLPGYMVPTRYGIVDTLPLTINGKLDVAAL
                     PEPIAATSGTGRTPRTDREATLLEIVASVLGIDGIGVDDDFFTLGGDSISSISVCGRA
                     RKAGLNITPRDVFRRRTIAALAASADTTQTPESGPDTGTGAITATPMLAETAQANTPL
                     TNFYQAMVLATPAGITAHEVQRVLQTVVDAHAMLRARLVTTATGWTLTVPDQTAAIGL
                     TVRSGALTSDCVEAEKAAAAAELDPRAGNMVRAVWFDAPEASGQLLLVIHHLVIDGVS
                     WRILSEDLSRAWNDLAAGRDVAADAVPTSFHTWANALAQRTFDDESEYWADVLATPDP
                     DLGSRPPDPAIDTADTVRSHEVTLASDVTEALLTSVPAAMHGGVNDVLLTGLAVALAQ
                     WRADRGHSQNTAAVISLEGHGRESDLVAADLDLSRTVGWFTSIYPVRIDPGPLEWDTV
                     RSAGAELAAAAKAVKEQLRAVPNRGLGYGVLRHLHGALDGTPPQILFNYLGRFTGGSG
                     NDWRPVAGIGALTEGVDPSNPAMSLEINALAEERPDGTVLSMTLAWPGGLLDADDVSE
                     LGSMWADVLVALTRCDALRGHTPSDFGLVTVSQDDIDGWDRLGEVEEVLPLLPLQEGM
                     YFHSMFGDPATDTYRVQQIAQLSGPVDPEVLRTSLTVVMRRHQALRASFNELTDGRVA
                     QVIWSDVPVQLTVVDTDDLESIAAEELARPFNLAEAPLVRYMLVRLGDDDHRLVQTMH
                     HIIADGWSYPVLFGDIVDHYNAAIGVGSAPQPITVTLRDHIETVTDRDRGAARQAWEQ
                     ALAGAEPTALVPRPDGAPVGEHRSVVRRLDSARTPAVGRAARVHGVTVGTVLHGAWGL
                     MLGRLLGRNRVVFGSTVSGRGGDLAGTESIVGLLINTIPVPMSWEPHATLASALIDLQ
                     DQQSALLDAQQMGLAELARLAGVREFFDTMVIVENFPSTSSAEHTDPRAVAFRGFTGT
                     DSPHYPVSFVAYADDQLTVEIKYDAGVVTPEQAERYAERVERILIAFAETPDLPVSRI
                     DLRTNAERQFTAGHASRPGPGRTLGASFAEVAATYPDAVAVSCGDTRLTYRELDDRAA
                     AVAATLAERGVGAESRVAIALPRSADLIVAVLAVIKAGGTYVPIDIGAPAARVQHILA
                     DSAPVCLLTDTAERFTGVPHVILAEAAQNPARPQAPTVSPDHAAYVIYTSGSTGVPKG
                     VEVTHRNVAALFAGTTSGLYDFGPDDVWTMFHSAAFDFSVWELWGPLLHGGRLVVVEH
                     DVARDPERFVDLLARERVTVLNQTPSAFYPLLEADARLRRQLALRYVIFGGEALDVRR
                     LAPWYANHESHSPRLVNMYGITETCVHVSHRALDTADTGAAGSVIGGPLPGLRIHLLD
                     NNLQPVPAGVVGEMYIAGGQVARGYTGRPGLTATRFVANPFDGAGERLYRSGDLAMWT
                     DAGELVYLGRSDAQVKVRGYRIELGEVEAALVTLPGVTNAAADVRHDDTGRARLIGYV
                     VGDALDIGALRSTLAERLPDYMVPSVLLRLDVLPLTVNGKLDRAALPDPEPVEQAPAP
                     VAGTGTASLLAGLCTEILGTTVGVEDDFFTAGGDSIIAIQLVNRARREGVRITPQQVF
                     VHRTPAALASVLDTGSVAAPDATPDEDAGPDLGEVMLTPIVQRLAELGGTVTRFNQSE
                     LLRTPAGATVARLETAINAVIARHDALRIRLHRPAPMLWSLETTAAAPVSITRIDAHG
                     FDDEQLRTVIATESDAAADRLDPENGVVVVAVWFDRGTDTQGRLLLVVHHLAVDGVSW
                     RILLDDLAEAYRQALSGQRVAPAPVTTSLRQYARTVNENAQHSSRLTEFEHWTEVLAP
```

Query Match 2.8%; Score 42.6; DB 2; Length 30683;
Best Local Similarity 54.9%; Pred. No. 0.35;
Matches 84; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

```
Qy    657 CACCTGGCGGTCGCTGACTGGCTGACGTCTGCACGGGAGGGGGAAGCGGCCCAGGCAGAC 716
          | ||||| | |||| ||| |  |  |  | ||| | ||  | ||| |||| | |
Db   4026 CGCCTGGTGCTCGCCGACCCCGAGGTGGTCGTTCTGGACGAGGCGACCGCCGAGGCCGGC 4085
```

```
Qy    717 AGATGGCAGGCGTTCGGACTGGAAGATAACGCCGCCGCCTTCAGCCTGGTCCTGGACAGA 776
          ||    || ||   || |||||| ||   |||| | |||  || |  || || |  ||
Db   4086 AGTGCACACGCCGCCGAACTGGAGGACGCCGCCTCGGCCGCCACCGCGGGCCGGACCACG 4145

Qy    777 CTGCGTGAGACGGAAAACTTCAAAAAAGACGCG 809
          |||   |   || |     |||   | | ||||
Db   4146 CTGATCGTCGCGCACCGGCTCACGCAGGCCGCG 4178
```

RESULT 9
AF034152
LOCUS AF034152 15223 bp DNA BCT 18-AUG-1998
DEFINITION Mycobacterium smegmatis exochelin gene cluster, ExiT (exiT) and FxbB (fxbB) genes, complete cds; and FxbC (fxbC) gene, partial cds.
ACCESSION AF034152
NID g3421057
VERSION AF034152.1 GI:3421057
KEYWORDS .
SOURCE Mycobacterium smegmatis.
ORGANISM Mycobacterium smegmatis
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
REFERENCE 1 (bases 1 to 15223)
AUTHORS Zhu,W., Arceneaux,J.E., Beggs,M.L., Byers,B.R., Eisenach,K.D. and Lundrigan,M.D.
TITLE Exochelin genes in Mycobacterium smegmatis: identification of an ABC transporter and two non-ribosomal peptide synthetase genes
JOURNAL Mol. Microbiol. 29 (2), 629-639 (1998)
MEDLINE 98385832
REFERENCE 2 (bases 1 to 15223)
AUTHORS Zhu,W.M., Arceneaux,J.E.L., Beggs,M.L., Byers,B.R., Eisenach,K.D. and Lundrigan,M.D.
TITLE Direct Submission
JOURNAL Submitted (11-NOV-1997) Pathology, University of Arkansas Medical Sciences Campus, 4301 W. Markham, Little Rock, AR 72205, USA
FEATURES Location/Qualifiers

```
     source          1. .15223
                     /organism="Mycobacterium smegmatis"
                     /strain="LR222"
                     /db_xref="taxon:1772"
                     /note="strain LR222 is a highly transformable strain
                     created by J. Crawford and M. Beggs at UAMS"
     gene            481. .3857
                     /gene="exiT"
     RBS             486. .489
                     /gene="exiT"
     mRNA            <489. .>3857
                     /gene="exiT"
     CDS             489. .3857
                     /gene="exiT"
                     /note="ABC Transporter; This ABC Transporter probably
                     excretes exochelin"
                     /codon_start=1
                     /transl_table=11
                     /product="ExiT"
                     /protein_id="AAC32046.1"
                     /db_xref="PID:g3421058"
                     /db_xref="GI:3421058"
                     /translation="MSVPVLIGIIVDRAVATGSVEAILRWIAVLAALFVVLTVVYRFG
                     ARLLMFAIARESHLLRVESSAKILDPLGIETDYKVGELLSISSDDADEVSYLLDYVPR
                     IVGAVVGTVVCGAVLLSIHLPLGLMVLIGVPVVVFGLQLTAPMIARRVEDQQAEIGRA
                     TALATDLISGQRPLQGNRRAGQRGRALPCREPSRADRDVACRAHPEHPLGRCGGGRGT
                     RRDGGGRGRGLLRDPRFDDGGPVDHRDRPRPVPHRTVLAAGGRAELGSRGPGVGQPRR
                     SRPERPDPPLPRLRGGPPRRRGPPVGPQRNTRGSEKPVVRRRARRIRRGTHERRPRCR
                     RTHRSALRISASRGQRHGAGRWPAPRPDPAERAPRTPARRTPPQLVVQRDAGNQSECH
                     RNRSAARTGLRRRRTTGVVRRRRRPAPRRSRPCRRGARCEPVRRTTPTLDPGPCAAG
                     RSGRPGAARSDHRDRHRHRAGHRRRHTPAAHARGTHNGGPHQQSGPARRGRPGAARHR
                     RASGQRRHTPRSGRRPLRLPGSGDPMISDEPVTPPVLPVATARRSGAWLIADLRGRRT
                     ALAAVVTVGLAAAAASVVPIYLLGMLVDRVEHGDGTRGLVALGTVIALAAVAGGLGTG
                     LSTYLTSRLGEQMLADLRERVLERALNLPATLIEESGRGDLLSRVGPDVAVVARTVAQ
                     VLPVILNGFFLGVVTLVGMTTLDWRLGLAGALAIPAYIAGLHWYLPRSAPMYADERIA
                     IANRAQVTVESIQGARTIDAYEIHDRHLADIDRASGRARDISIAVFTFFTRLVGRVNR
                     AEFIGLAATLVVGFLLVRSGETTVGQVTAAALMFHRLFNPVGELMYNFDEIQSASASL
                     ARLVGVIDLLDAPGPDARRTAPARPHGADISVENVSFGYGTGPEVVHGVSLHVPAGTRC
                     ALVGTTGAGKTTLAAIVAGLLTPTSGHARIGGVPVTEIADLRRWVATITQEVHVFSGP
                     LIDDLRLVAPQASVDDVWTALETVGASTWVRALDDGLDTVVGEHGHELTAAQAQHVAM
                     ARLVLADPEVVVLDEATAEAGSAHAAELEDAASAATAGRTTLIVAHRLTQAAQADQVA
                     VMADGRIVENGTHAELIALGGRYARLWQAWIRHN"
misc_feature         3226. .3277
                     /gene="exiT"
                     /note="encodes ATP/GTP-binding site motif A (P-loop)"
gene                 3876. .8810
                     /gene="fxbB"
RBS                  3876. .3881
                     /gene="fxbB"
mRNA                 <3891. .>8810
                     /gene="fxbB"
CDS                  3891. .8810
                     /gene="fxbB"
                     /function="involved in assembly of exochelin molecule"
                     /note="peptide synthetase"
                     /codon_start=1
                     /transl_table=11
                     /product="FxbB"
                     /protein_id="AAC32047.1"
                     /db_xref="PID:g3421059"
                     /db_xref="GI:3421059"
                     /translation="MTADSLDIAELLELWNNHSTPERTSTVPTLFAAQCALTPDEVAV
                     VDGERRLTYRHLETHVAQLAHAVRVAAGEGPEPIVAIGVPRSAEMVVCVLAAMMAGVA
                     FVPLDPSWPAHRRRQVLADSGAVATFITREDESDWGVPGLRVDLGAWQFTAESPVLPQ
                     ADVHPAQLAYVIFTSGSTGKPKGAMIRHDAIAERLQWQRDHILHFGKHDHTDASLFKA
                     PLSFDISVNEILLPLVSGGRVVVAVPDGEKDPEYLLELIRTEQVTFVYLVSSMLDTLL
                     ELDRLATADGAPSSLASLRHVWCGGEVLTPGLFARFRKQLTTTLYHGYGPAEATIGVS
                     HVIYRDTAERIATSIGRPNPHTQLYVLDEYLRPVPPGVGGELYAAGFLLGRGYVNAPS
                     LTASRFVANPFDGNGSRMYRTGDLARWTEDGSLEFLGRADNQVKIGGRRVELEEIESQ
                     LADHPAVRHAVVDVHRQGGADVLVGYLVAADGVRNDAAWHAEVADWARTRLPEYMVPK
                     AFVALDQVPLTANGKTDRRALPAPDLERSGTVKPPRTPRETVLCQVFADALDIDAVGV
                     DEDFFALGGDSIVAIRVVSRLRAAGYTLRPRDMFAHRTVEALAPLLGDSDVRDTGPAV
                     DPTGAATPTPILRWLDEVGTAGSVLNGFHQGMSLVTPADADENTLRAAIAATVRRHHV
                     LWAPPGRTASDIDIPGTPPENRLLIADASDGIPAQAEKVARQLVSLLDTARIAFGWIR
                     RPAAPGRLVVIADHTVIDGVSLRILAEDIATAYGLIAEGRAVELPTPHTSWRAWAQRL
                     ADTAAAGGFDADLEHWQQVCATTETPWGDRALDPAIDTVATESRLTVELPSAVTDAIL
                     TTVPDRIHRHINDALVAALYLALRRWLHTRGVGADTLLVEMEGHGREGHLVDSDTAGL
                     DLSNTVGWFTTLYPVALRDAEFDWQAAVSGGPQLGAAVRSVKDQLRSVPSHGFSYGAL
                     RYLRDGTSGLEAAPQVLFNYLGRFGTADRPWALANDTTAVLEDRDPGMPLPRLLEVNA
                     EAVTTADGSVLRATFSWPAHAVAEVDVRTLAGMWTDLLTAIASSDDVRGHSASDFDRV
                     SVTADDVAELERRYPGLTDLLPLTPTQQGIYFHSTFSRDRDPYVVQQIVDITGPLDTE
                     RFQRATESVAARHRALGAVFTTLSDGTPVAVHAATVAPDFDVLDARHAADPSTVVAQR
                     ARWERERRFDLAAAPLTRYTLVRRRDDLHTMIETVHHIVADGWSVPIVLDDLLTAYAG
                     DDFDGPAPEFARFVDWLEQRDTDSDRAAWAPALEDITEPTRLAAADGARGRTTGSGFG
                     TRTVTLQSRSAVADAANGAAVTVGTLLHTAWGLALGRLTGRDDVVFGTVVSGRGADVH
                     GIESMVGLLVNTVPVRVAWSADDTGADVAGRLAAVESTLLEHHHLPLTEAHRLAGVGE
                     LFDTLVVIENLGATTHTRGDLTLGDIGVIEAPHYPLTLMISVRDTITVTVTNDREHVS
                     DVLADTAVAAFTEVLTALTADPGVSPGDIALPAPQATAPPTQAPQTVTDLIAAAIAEH
                     RDDIALVVGETEWTYGNSAPAQVNSPHRWPRRVCAAATSWRSRPFVRRTWWPRSGRSS
                     PRARRISRSTSPIRGPVSSTCCAMRAPPR"
misc_feature         4387. .4448
                     /gene="fxbB"
                     /note="encodes putative AMP-binding domain signature"
misc_feature         5589. .5666
                     /gene="fxbB"
                     /note="encodes phosphopantetheine attachment site"
misc_feature         7965. .8081
                     /gene="fxbB"
                     /note="encodes binding-protein-dependent transport systems inner membrane component signature"
RBS                  8824. .8828
                     /gene="fxbC"
gene                 8824. .>15223
                     /gene="fxbC"
mRNA                 <8837. .>15223
                     /gene="fxbC"
CDS                  8837. .>15223
                     /gene="fxbC"
```

```
                     /function="involved in assembly of exochelin molecule"
                     /note="peptide synthetase"
                     /codon_start=1
                     /transl_table=11
                     /product="FxbC"
                     /protein_id="AAC32048.1"
                     /db_xref="PID:g3421060"
                     /db_xref="GI:3421060"
                     /translation="MSGALPADTIVVSTTATHAAVPFTPVPVDGADAVSVLYTSGSTG
                     EPKAVVGTHAALANRLAWAVEAWPAATRIAKSPLSFIDGTTELLAGLAAGARTVLAGD
                     ETARDGRRLAQLVAAHGVEQLLAVPSLAAVLADERTEDVAELNRWIVSGEALEPRHLH
                     ALRTACPTAEIVNSYGSSEVAGDVLAGVQDDAGITLGAAVPGAGIRILDSRLRQLPAG
                     VIGEIYVTGGQLARGYLGRPGQTATRFVAAPGGERMYRTGDLGALLPGGRVVFAGRAD
                     DQLKINGHRVEPGEIESVLARQPGVREAAVIGTGTQLAAFVVLESDAPGAGDLLTAVS
                     AELPGHLVPSSLRPVDAIPLLPNGKRDNNALRSLLVPGESTGTAPTNDVERAVLDVMT
                     GVLGRDPLGADDDFFAVGGDSISAIRVTSRLARAGYHIATEDVFRGRTAAGVAALVDT
                     VESLAHDAAIEPFATVRLSPQTIDRIRESGQVEDIWAMSPLQLGVYYQSTLADAGPTY
                     IAQNVFEFDRRIDVDAMRRAFTALLRRHPQLRAGFRTVEHAEPEPAPDATPVVQVVVT
                     DPPSDLTVVDLTDQDDPAAAQHIVDTDRTAPFDVATPPLLRLTVIRLPGGRDRMLFTY
                     HFLLFDGWSRELVLRDLFALYDSDAQHGAIAPHGDLVVRHLQWLATDGDPQASGARDA
                     WRDLLAGLTEPTLASGVSPDHPDARPGTEPGRIVVRVPEDVTTQLHTCATAHGATLNS
                     VVTAAVALVTGYHAGTTDVVIGTTVAGRPGHLVGIDETIGLFLNTVPVRVDMSPTRSV
                     AAAMASIGEQRVAMMRHDHLGLGQIQRAAGDSGSALFDSLLVLQNFLDDDTFTDLESR
                     HGIVDVSYHDTTHFPLTWVLTPGRELAIKLEHRVVDDARATEMVEQLVTVLRTIAAEP
                     DTAVGAVDLIGADRRAQLERRWSSTERPLEPVTIAELLARRAEQNPDDVALVFGAQSV
                     TYREFDDRVSQFARHLRARGAAPETFVALALPRSIDMVVALFAVLRAGAAYLPLELDL
                     PIDRLRTIIDNAEPVLLVTTTDRTELIGHARARGADVIALDDAETAATLADTPAHPLT
                     AGELGAFASDSTRLNHPAYLIYTSGSTGRPKGVLTGYAGLTNMYFNHREAIFAPTVAR
                     AGSAEQLRIAHTVSFSFDMSWEELFWLVEGHQVHVCDEELRRDAPALVAYCHRHRIDV
                     INVTPTYAHHLFDAGLLDDGAHTPPLVLLGGEAVGDGVWSALRDHPDSAGYNLYGPTE
                     YTINTLGGGTDDSDTPTVGQPIWNTRGYILDAALRPVPDGAVGELYIAGTGLALGYHR
                     RAGLTAATMVADPYVPGGRMYRTGDLVRRRPGSAELLDYLGRVDDQVKIRGYRVELGE
                     IESVLTRADGVARCAVVARATGANPPVKTLAAYVIPDRWPAEDAAFITGLRDHLARVL
                     PGYMVPTRYGIVDTLPLTINGKLDVAALPEPIAATSGTGRTPRTDREATLLEIVASVL
                     GIDGIGVDDDFFTLGGDSISSISVCGRARKAGLNITPRDVFRRRTIAALAASADTTQT
                     PESGPDTGTGAITATPMLAETAQANTPLTNFYQAMVLATPAGITADEVQRVLQTVVDA
                     HAMLRARLVTTATGWTLTVPDQTAAIGLTVRSGALTSDCVEAEKAAAAAELDPRAGNM
                     VRAVWFDAPEASGQLLLVIHHLVIDGVSWRILSEDLSRAWNDLAAGRDVAADAVPTSF
                     HTWANALAQRTFDDESEYWADVLATPDPDLGSRPPDPAIDTADTVRSHEVTLASDVTE
                     ALLTSVPAAMHGGVNDVLLTGLAVALAQWRADRGHSQNTAAVISLEGHGRESDLVAAD
                     LDLSRTVGWFTSIYPVRIDPGPLEWDTVRSAGAELAAAAKAVKEQLRAVPNRGLGYGV
                     LRHLHGALDGTPPQILFNYLGRFTGGSGNDWRPVAGIGALTEGVDPSNPAMSLEINAL
                     AEERPDGTVLSMTLAWPGGLLDADDVSELGSMWADVLVALTRCDALRGHTPSDFGLVT
                     VSQDDIDGWDRLGEVEEVLPLLPLQEGMYFHSMFGDPATDTYRVQQIAQLSGPVDPEV
                     LRTSLTVVMRRHQALRASFNELTDGRVAQVIWSDVPVQLTVVDTDDLESIAAEEL"
     misc_feature    8927. .8992
                     /gene="fxbC"
                     /note="encodes putative AMP-binding domain signature"
     misc_feature    9653. .9708
                     /gene="fxbC"
                     /note="encodes ATP-dependent DNA ligase signatures"
     misc_feature    10043. .10120
                     /gene="fxbC"
                     /note="encodes phosphopantetheine attachment site"
     misc_feature    12143. .12208
                     /gene="fxbC"
                     /note="encodes putative AMP-binding domain signature"
BASE COUNT     2329 a   5576 c   5039 g   2279 t
ORIGIN


  Query Match             2.8%;  Score 42.6;  DB 2;  Length 15223;
  Best Local Similarity   54.9%;  Pred. No. 0.33;
  Matches   84;  Conservative    0;  Mismatches   69;  Indels    0;  Gaps    0;

Qy     657 CACCTGGCGGTCGCTGACTGGCTGACGTCTGCACGGGAGGGGGAAGCGGCCCAGGCAGAC 716
           | ||||| | |||| |||       |  |   | | ||| | ||   | ||| |||| | |
Db    3582 CGCCTGGTGCTCGCCGACCCCGAGGTGGTCGTTCTGGACGAGGCGACCGCCGAGGCCGGC 3641

Qy     717 AGATGGCAGGCGTTCGGACTGGAAGATAACGCCGCCGCCTTCAGCCTGGTCCTGGACAGA 776
           ||    || ||   || |||||| ||   |||| | |||  || |  || || |  ||
Db    3642 AGTGCACACGCCGCCGAACTGGAGGACGCCGCCTCGGCCGCCACCGCGGGCCGGACCACG 3701

Qy     777 CTGCGTGAGACGGAAAACTTCAAAAAAGACGCG 809
           |||   |   || |     |||   | | ||||
Db    3702 CTGATCGTCGCGCACCGGCTCACGCAGGCCGCG 3734


RESULT  10
BACLEVRA
LOCUS       BACLEVRA     2935 bp    DNA             BCT       03-JAN-1994
DEFINITION  Bacillus subtilis levR gene, complete cds.
ACCESSION   M60105
NID         g143149
VERSION     M60105.1  GI:143149
KEYWORDS    levR gene; transcription regulator.
SOURCE      B.subtilis (168 Marburg) DNA.
  ORGANISM  Bacillus subtilis
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Bacillus.
REFERENCE   1  (bases 1 to 2935)
  AUTHORS   Debarbouille,M., Martin-Verstraete,I., Klier,A. and Rapoport,G.
  TITLE     The transcriptional regulator LEVR of Bacillus subtilis has domains
            homologous to both sigma 54 and PTS dependent regulators
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 88, 2212-2216 (1991)
  MEDLINE   91172782
FEATURES             Location/Qualifiers
     source          1. .2935
                     /organism="Bacillus subtilis"
                     /db_xref="taxon:1423"
     gene            49. .2920
                     /gene="levR"
     CDS             49. .2865
                     /gene="levR"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA22572.1"
                     /db_xref="PID:g143150"
                     /db_xref="GI:143150"
                     /translation="MIDVRRIDKIYHQLKHNFHDSTLDHLLKIQGNSAKEIAEQLKME
                     RSNVSFELNNLVRSKKVIKIKTFPVRYIPVEIAEKLFNKKWDTEMMEVKDLQAFSGNS
                     KQNHQHISTNPLELMIGAKGSLKKAISQAKAAVFYPPNGLHMLLLGPTGSGKSLFANR
                     IYQFAIYSDILKAGAPFITFNCADYYNNPQLLLSQLFGHKKGSFTGAAEDKAGLVEQA
                     NGGILFMDEIHRLPPEGQEMLFYFIDSGSYNRLGESEHKRTSNVLFICATTENPSSAL
                     LKTFLRRIPMTIHIPSLEERSLNERVDLTTFLLGKEAERIKKNLSVHIDVYNALIHSA
                     KFGNVGQLKSNVQLVCAHGFLHNLDRNEVIELTVRDLPDEIKQEWMSSSKNMQRSKAI
                     SEYVNITTIISPIVEDETTKIDEDLSFNLYHLIEEKVKTLMKEGLSKKDINQYILTDV
                     HLHVRSFFHHQAFQKDNLLTFVEDDVIQMTKQLKEIAEHELDCTFDRKFIYFLSMHID
                     AFLKRGKQIDVLNTQETDEIRDTHVKEYRVAMIFKDKIQEYFKVAIPEIEVIYLTMLI
                     HSIKSLKENKRVGIIVAAHGNSTASSMVEVATELLGSTPIAAVDMPLTVSPSDILECV
                     AEKMKQVDEGEGVLMLVDMGSLAMLESRLEEKTGISIKTISNVTTSMVLDAVRKVNYL
                     NLNLHAIYQSVTKDFIELWERQPAASGKKKALVSICTTGSGTAKKLEDILTTIVNKAS
                     DTPIHILTVSSIKLANSIKEIEKEYEILATVGTKDPKINAPHVSLEVLIEGEGEKLIQ
                     QAITKGSISLSNGLNEANIIVRELCEDSLKKYLVFLNPHHVIDMLLEWLQTVQDELGV
                     IFNNAVLIKVIMHTAFAFERVIKQNPIAFLEEEEINDQLKEMVYVTERTLAPYEEKLG
                     LRISDDEKLFIAAIFAEEVHGQLF"
     misc_structure  2883. .2920
                     /gene="levR"
                     /note="palindromic structure"
BASE COUNT      963 a    534 c    644 g    794 t
ORIGIN


  Query Match             2.7%;  Score 41.6;  DB 1;  Length 2935;
  Best Local Similarity   46.3%;  Pred. No. 0.55;
  Matches  137;  Conservative    0;  Mismatches  159;  Indels    0;  Gaps    0;

Qy     850 CTGAAGATGCTGCGCTGAGAGCAAAAACCTTTGCCATGGCAACAGAGGCAACATCAACCT 909
           ||||| |||       | | | | |  | || | | || | |  ||| | |||| |||
Db    1226 CTGAATATGTCAATATTACAACGATTATCTCTCCGATTGTAGAAGATGAAACAACAAAAA 1285

Qy     910 GCGAGGACCGGGTCACACATGCCCTGCACCAGATGAATAACGTACAACTGGTACATAATG 969
             || ||     || ||  |  | |  | ||  ||| | | | | || ||  |     |
Db    1286 TTGACGAAGACCTCTCATTTAACTTATATCATTTGATTGAGGAAAAAGTGAAAACGCTTA 1345

Qy     970 CAGAAAAAGGGGGAATACGACAACAATCTCCAGGGGCTGGTTTCCACGGGGCGTGAGATGT 1029
```

```
Db    1346 TGAAAGAAGGTTTATCCAAAAAGGATATTAACCAGTACATCTTGACTGATGTCCATTTGC 1405

Qy    1030 TCCGCCTGGCAACACTGGAACAGATTGCCCGGGAAAAAGCCGGAACACTGGCTTTAGTCG 1089

Db    1406 ATGTCAGAAGTTTTTTTCATCATCAAGCTTTTCAAAAAGACAACCTGCTGACATTCGTTG 1465

Qy    1090 ATGACGTTGAGGTCTATCTGGCGTTCCAGAATAAGCTGAAGGAATCACTTGAGCTG 1145

Db    1466 AAGACGATGTGATTCAAATGACAAAACAGTTAAAAGAAATAGCTGAACATGAGCTG 1521


RESULT  11
BS233DEG
LOCUS       BS233DEG    23911 bp    DNA             BCT       19-MAY-1997
DEFINITION  B.subtilis 23.9kb fragment from map position 233 degrees on the
            chromosome.
ACCESSION   X92868 X79978
NID         g2108259
VERSION     X92868.1  GI:2108259
KEYWORDS    aapA gene; adhA gene; adhB gene; alcohol dehydrogenase; amino acid
            permease; chitosanase; csn gene; fructose-specific IIA component;
            fructose-specific IIB component; fructose-specific IIC component;
            fructose-specific IID component; levanase; levD gene; levE gene;
            levF gene; levG gene; levR gene; mercuric resistance;
            NADP-dependent; sacC gene; spore coat protein F; transcriptional
            regulatory protein; yraA gene; yraB gene; yraC gene; yraD gene;
            yraE gene; yraF gene; yraG gene; yraH gene; yraI gene; yraJ gene;
            yraK gene; yraL gene; yraM gene; yraN gene; yraO gene.
SOURCE      Bacillus subtilis.
  ORGANISM  Bacillus subtilis
            Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillaceae;
            Bacillus.
REFERENCE   1  (bases 1 to 23911)
  AUTHORS   Parro,V., San Roman,M., Galindo,I., Purnelle,B., Bolotin,A.,
            Sorokin,A. and Mellado,R.P.
  TITLE     A 23911 bp region of the Bacillus subtilis genome comprising genes
            located upstream and downstream of the lev operon
  JOURNAL   Microbiology 143 (Pt 4), 1321-1326 (1997)
  MEDLINE   97286548
REFERENCE   2  (bases 1 to 23911)
  AUTHORS   Mellado,R.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-NOV-1995) R.P. Mellado, Centro Nacional de
            Biotecnologa (CSIC), Campus de la Universidad Autnoma, Cantoblanco
            28049 Madrid, E - 28049 Madrid, SPAIN
  REMARK    Revised by submittor 10-OCT-96
COMMENT     Overlaps with X79978, X56098, X05649 and M60105.
FEATURES             Location/Qualifiers
     source          1. .23911
                     /organism="Bacillus subtilis"
                     /strain="168"
                     /db_xref="taxon:1423"
                     /map="233 degrees"
     gene            complement(376. .1755)
                     /gene="aapA"
     CDS             complement(376. .1755)
                     /gene="aapA"
                     /codon_start=1
                     /transl_table=11
                     /product="amino acid permease"
                     /protein_id="CAA63459.1"
                     /db_xref="PID:e209877"
                     /db_xref="PID:g2108260"
                     /db_xref="GI:2108260"
                     /db_xref="SWISS-PROT:O06005"
                     /translation="MIGNSSKDNFGQQQKLSRGLKNRHIQLMAIGGAIGTGLFLGSGK
                     SIHFAGPSILFAYLITGVFCFFIIRSLGELLLSNAGYHSFVDFVRDYLGNMAAFITGW
                     TYWFCWISLAMADLTAVGIYTQYWLPDVPQWLPGLLALIILLIMNLATVKLFGELEFW
                     FALIKVIAILALIVTGILLIAKGFSAASGPASLNNLWSHGGMFPNGWHGFILSFQMVV
                     FAFVGIELVGLTAGETENPQKVIPKAINQIPVRILLFYVGALFVIMCIYPWNVLNPNE
                     SPFVQVFSAVGIVVAASLINFVVLTSAASAANSALFSTSRMVYSLAKDHHAPGLLKKL
                     TSSNVPSNALFFSSIAILIGVSLNYLMPEQVFTLITSVSTICFIFIWGITVICHLKYR
                     KTRQHEAKANKFKMPFYPLSNYLTLAFLAFILVILALANDTRIALFVTPVWFVLLIIL
                     YKVQTQQRT"
RBS                  complement(1765. .1769)
RBS                  2461. .2465
gene                 2472. .5288
                     /gene="levR"
CDS                  2472. .5288
                     /gene="levR"
                     /codon_start=1
                     /transl_table=11
                     /product="transcriptional regulatory protein"
                     /protein_id="CAA63460.1"
                     /db_xref="PID:e209878"
                     /db_xref="PID:g2108261"
                     /db_xref="GI:2108261"
                     /db_xref="SWISS-PROT:P23914"
                     /translation="MIDVRRIDKIYHQLKHNFHDSTLDHLLKIQGNSAKEIAEQLKME
                     RSNVSFELNNLVRSKKVIKIKTFPVRYIPVEIAEKLFNKKWDTEMMEVKDLQAFSGNS
                     KQNHQHISTNPLELMIGAKGSLKKAISQAKAAVFYPPNGLHMLLLGPTGSGKSLFANR
                     IYQFAIYSDILKAGAPFITFNCADYYNNPQLLLSQLFGHKKGSFTGAAEDKAGLVEQA
                     NGGILFMDEIHRLPPEGQEMLFYFIDSGSYNRLGESEHKRTSNVLFICATTENPSSAL
                     LKTFLRRIPMTIHIPSLEERSLNERVDLTTFLLGKEAERIKKNLSVHIDVYNALIHSA
                     KFGNVGQLKSNVQLVCAHGFLHNLDRNEVIELTVRDLPDEIKQEWMSSSKNMQRSKAI
                     SEYVNITTIISPIVEDETTKIDEDLSFNLYHLIEEKVKTLMKEGLSKKDINQYILTDV
                     HLHVRSFFHHQAFQKDNLLTFVEDDVIQMTKQLKEIAEHELDCIFDRKFIYFLSMHID
                     AFLKRGKQIDVLNTQETDEIRDTHVKEYRVAMIFKDKIQEYFKVAIPEIEVIYLTMLI
                     HSIKSLKENKRVGIIVAAHGNSTASSMVEVATELLGSTPIAAVDMPLTVSPSDILECV
                     AEKMKQVDEGEGVLMLVDMGSLAMLESRLEEKTGISIKTISNVTTSMVLDAVRKVNYL
                     NLNLHAIYQSVTKDFIELWERQPAASGKKKALVSICTTGSGTAKKLEDILTTIVNKAS
                     DTPIHILTVSSIKLANSIKEIEKEYEILATVGTKDPKINAPHVSLEVLIEGEGEKLIQ
                     QAITKGSISLSNGLNEANIIVRELCEDSLKKYLVFLNPHHVIDMLLEWLQTVQDELGV
                     IFNNAVLIKVIMHTAFAFERVIKQNPIAFLEEEEINDQLKEMVYVTERTLAPYEEKLG
                     LRISDDEKLFIAAIFAEEVHGQLF"
RBS                  5464. .5468
gene                 5478. .5918
                     /gene="levD"
CDS                  5478. .5918
                     /gene="levD"
                     /codon_start=1
                     /transl_table=11
                     /product="fructose-specific IIA component"
                     /protein_id="CAA63461.1"
                     /db_xref="PID:e209879"
                     /db_xref="PID:g2108262"
                     /db_xref="GI:2108262"
                     /db_xref="SWISS-PROT:P26379"
                     /translation="MISVIISGHGDFPIALKESSGMIFGEENNLIAVPFFKGEGIQTL
                     QEKYHQALKDIPEEHEVLFLVDIFGGTPYNAAASFIAEDQRMDMAAGVNLPILLEVLS
                     LREHLALKDLLNNLKAMSQQSFQVCSEHLEKVKTANQDTREDEL"
RBS                  5904. .5907
                     /gene="levD"
gene                 5918. .6406
                     /gene="levE"
CDS                  5918. .6406
                     /gene="levE"
                     /codon_start=1
                     /transl_table=11
                     /product="fructose-specific IIB component"
                     /protein_id="CAA63462.1"
                     /db_xref="PID:e209880"
                     /db_xref="PID:g2108263"
                     /db_xref="GI:2108263"
                     /translation="MNIVLARIDDRFIHGQILTRWIKVHAADRIIVVSDDIAQDEMRK
                     TLILSVAPSNVKASAVSVSKMAKAFHSPRYEGVTAMLLFENPSDIVSLIEAGVPIKTV
                     NVGGMRFENHRRQITKSVSVTEQDIKAFETLSDKGVKLELRQLPSDASEDFVQILRNV
                     TK"
RBS                  6409. .6417
gene                 6423. .7232
                     /gene="levF"
CDS                  6423. .7232
                     /gene="levF"
                     /codon_start=1
                     /transl_table=11
```

```
                     /product="fructose-specific IIC component"
                     /protein_id="CAA63463.1"
                     /db_xref="PID:e209881"
                     /db_xref="PID:g2108264"
                     /db_xref="GI:2108264"
                     /db_xref="SWISS-PROT:P26381"
                     /translation="MSSLQIILLLIIAAITGIASVLDEGQTHRPLVACTLVGLVLGDL
                     KTGIILGGTLELMALGWMNVGLAMAPDTAIASVISTILVITADQGIGEGIAVAVALAA
                     AGQALTIFVRTITVFFIHRADQYAKAGNIKGIEIMHITAMVFQALRVMIPTLIVALIS
                     VSAVQAFLGNIPDVITKGLQIGGGIIVVVGYAMVINMMNIPYLKPFFYIGFLLAAFTD
                     FNLVGFGALGLCLALLYQQVMQKQSAHGAVAAASDSGSVAVYDDDDDDLDA"
RBS                  7239. .7244
gene                 7253. .8080
                     /gene="levG"
CDS                  7253. .8080
                     /gene="levG"
                     /codon_start=1
                     /transl_table=11
                     /product="fructose-specific IID component"
                     /protein_id="CAA63464.1"
                     /db_xref="PID:e209882"
                     /db_xref="PID:g2108265"
                     /db_xref="GI:2108265"
                     /db_xref="SWISS-PROT:P26382"
                     /translation="MEKEKRLTKKEIFSMFIRSNFLLGSFNFERVQAMGYCYVMIPAI
                     KKLYGPGAKRNEALQRHLEWFNTHPWLTAPIFGVTAAMEEEMANNKGIDGKAISGMKI
                     GLMGPIAGVGDPIFWGTIRPVLAALGASLALGGNIAGPLLFFFLLNAIRLSTKYYGLK
                     YGYVKGMEILQDLAGNRIQKLTEGASILGLFVMGALVSKWTTINIPIVVSRIKDESGK
                     VDVQTVQNVLDSIMPGALPLGLTLLVAWMLRKGVNPLLIICGIFVIGILGYWAGFLA"
RBS                  8221. .8225
gene                 8237. .10270
                     /gene="sacC"
CDS                  8237. .10270
                     /gene="sacC"
                     /codon_start=1
                     /transl_table=11
                     /product="levanase precursor"
                     /protein_id="CAA63465.1"
                     /db_xref="PID:e209883"
                     /db_xref="PID:g2108266"
                     /db_xref="GI:2108266"
                     /db_xref="SWISS-PROT:P05656"
                     /translation="MKKRLIQVMIMFTLLLTMAFSADAADSSYYDEDYRPQYHFTPEA
                     NWMNDPNGMVYYAGEYHLFYQYHPYGLQWGPMHWGHAVSKDLVTWEHLPVALYPDEKG
                     TIFSGSAVVDKNNTSGFQTGKEKPLVAIYTQDREGHQVQSIAYSNDKGRTWTKYAGNP
                     VIPNPGKKDFRDPKVFWYEKEKKWVMVLAAGDRILIYTSKNLKQWTYASEFGQDQGSH
                     GGVWECPDLFELPVDGNPNQKKWVMQVSVGNGAVSGGSGMQYFVGDFDGTHFKNENPP
                     NKVLWTDYGRDFYAAVSWSDIPSTDSRRLWLGWMSNWQYANDVPTSPWRSATSIPREL
                     KLKAFTEGVRVVQTPVKELETIRGTSKKWKNLTISPASHNVLAGQSGDAYEINAEFKV
                     SPGSAAEFGFKVRTGENQFTKVGYDRRNAKLFVDRSESGNDTFNPAFNTGKETAPLKP
                     VNGKVKLRIFVDRSSVEVFGNDGKQVITDIILPDRSSKGLELYAANGGVKVKSLTIHP
                     LKKVWGTTPFMSNMTGWTTVNGTWADTIEGKQGRSDGDSFILSSASGSDFTYESDITI
                     KDGNGRGAGALMFRSDKDAKNGYLANVDAKHDLVKFFKFENGAASVIAEYKTPIDVNK
                     KYHLKTEAEGDRFKIYLDDRLVIDAHDSVFSEGQFGLNVWDATAVFQNVTKES"
gene                 complement(10358. .10822)
                     /gene="yraA"
CDS                  complement(10358. .10822)
                     /gene="yraA"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA63466.1"
                     /db_xref="PID:e250832"
                     /db_xref="PID:g2108267"
                     /db_xref="GI:2108267"
                     /db_xref="SWISS-PROT:O06006"
                     /translation="MSKKIAVLVTDQFEDIEYTSPVKAYEEAGYSVVAIDLEAGKEVT
                     GKHGEKVKIDKAISDVDASDFDALLIPGGFSPDLLRADDRPGEFAKAFVENKKPVFAI
                     CHGPQVLIDTDLLKGKDITGYRSIRKDLINAGANYKDAEVVVSHNIVDKQNT"

Query Match              2.7%;  Score 41.6;  DB 1;  Length 23911;
Best Local Similarity   46.3%;  Pred. No. 0.64;
Matches  137;  Conservative    0;  Mismatches  159;  Indels    0;  Gaps    0;

Qy     850 CTGAAGATGCTGCGCTGAGAGCAAAAACCTTTGCCATGGCAACAGAGGCAACATCAACCT 909
           ||||| |||       | | | | |  | || | | || | |  ||| | |||| |||
Db    3649 CTGAATATGTCAATATTACAACGATTATCTCTCCGATTGTAGAAGATGAAACAACAAAAA 3708

Qy     910 GCGAGGACCGGGTCACACATGCCCTGCACCAGATGAATAACGTACAACTGGTACATAATG 969
           || ||     || ||  |  | |  | ||  ||| | | | | || ||  |     |
Db    3709 TTGACGAAGACCTCTCATTTAACTTATATCATTTGATTGAGGAAAAAGTGAAAACGCTTA 3768

Qy     970 CAGAAAAAGGGGAATACGACAACAATCTCCAGGGGCTGGTTTCCACGGGGCGTGAGATGT 1029
           || ||||   || | | ||  || | |   |   | |  || |      |  ||
Db    3769 TGAAAGAAGGTTTATCCAAAAAGGATATTAACCAGTACATCTTGACTGATGTCCATTTGC 3828

Qy    1030 TCCGCCTGGCAACACTGGAACAGATTGCCCGGGAAAAAGCCGGAACACTGGCTTTAGTCG 1089
           |          |  | ||    ||     |||||| |      ||| | || | |
Db    3829 ATGTCAGAAGTTTTTTTCATCATCAAGCTTTTCAAAAAGACAACCTGCTGACATTCGTTG 3888

Qy    1090 ATGACGTTGAGGTCTATCTGGCGTTCCAGAATAAGCTGAAGGAATCACTTGAGCTG 1145
           | |||| || | |  |  || |    |||   ||    |  |    || |||||||
Db    3889 AAGACGATGTGATTCAAATGACAAAACAGTTAAAAGAAATAGCTGAACATGAGCTG 3944

RESULT  12
BSUB0014/c
LOCUS       BSUB0014   213420 bp    DNA             BCT       26-NOV-1997
DEFINITION  Bacillus subtilis complete genome (section 14 of 21): from 2599451
            to 2812870.
ACCESSION   Z99117 AL009126
NID         g2634966
VERSION     Z99117.1  GI:2634966
KEYWORDS    .
SOURCE      Bacillus subtilis.
  ORGANISM  Bacillus subtilis
            Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillaceae;
            Bacillus.
REFERENCE   1  (bases 1 to 213420)
  AUTHORS   Kunst,F., Ogasawara,N., Moszer,I., Albertini,A.M., Alloni,G.,
            Azevedo,V., Bertero,M.G., Bessieres,P., Bolotin,A., Borchert,S.,
            Borriss,R., Boursier,L., Brans,A., Braun,M., Brignell,S.C.,
            Bron,S., Brouillet,S., Bruschi,C.V., Caldwell,B., Capuano,V.,
            Carter,N.M., Choi,S.K., Codani,J.J., Connerton,I.F., Cummings,N.J.,
            Daniel,R.A., Denizot,F., Devine,K.M., Dusterhoft,A., Ehrlich,S.D.,
            Emmerson,P.T., Entian,K.D., Errington,J., Fabret,C., Ferrari,E.,
            Foulger,D., Fritz,C., Fujita,M., Fujita,Y., Fuma,S., Galizzi,A.,
            Galleron,N., Ghim,S.Y., Glaser,P., Goffeau,A., Golightly,E.J.,
            Grandi,G., Guiseppi,G., Guy,B.J., Haga,K., Haiech,J., Harwood,C.R.,
            Henaut,A., Hilbert,H., Holsappel,S., Hosono,S., Hullo,M.F.,
            Itaya,M., Jones,L., Joris,B., Karamata,D., Kasahara,Y.,
            Klaerr-Blanchard,M., Klein,C., Kobayashi,Y., Koetter,P.,
            Koningstein,G., Krogh,S., Kumano,M., Kurita,K., Lapidus,A.,
            Lardinois,S., Lauber,J., Lazarevic,V., Lee,S.M., Levine,A., Liu,H.,
            Masuda,S., Mauel,C., Medigue,C., Medina,N., Mellado,R.P.,
            Mizuno,M., Moestl,D., Nakai,S., Noback,M., Noone,D., O'Reilly,M.,
            Ogawa,K., Ogiwara,A., Oudega,B., Park,S.H., Parro,V., Pohl,T.M.,
            Portetelle,D., Porwollik,S., Prescott,A.M., Presecan,E., Pujic,P.,
            Purnelle,B., Rapoport,G., Rey,M., Reynolds,S., Rieger,M.,
            Rivolta,C., Rocha,E., Roche,B., Rose,M., Sadaie,Y., Sato,T.,
            Scanlan,E., Schleich,S., Schroeter,R., Scoffone,F., Sekiguchi,J.,
            Sekowska,A., Seror,S.J., Serror,P., Shin,B.S., Soldo,B.,
            Sorokin,A., Tacconi,E., Takagi,T., Takahashi,H., Takemaru,K.,
            Takeuchi,M., Tamakoshi,A., Tanaka,T., Terpstra,P., Tognoni,A.,
            Tosato,V., Uchiyama,S., Vandenbol,M., Vannier,F., Vassarotti,A.,
            Viari,A., Wambutt,R., Wedler,E., Wedler,H., Weitzenegger,T.,
            Winters,P., Wipat,A., Yamamoto,H., Yamane,K., Yasumoto,K., Yata,K.,
            Yoshida,K., Yoshikawa,H.F., Zumstein,E., Yoshikawa,H. and
            Danchin,A.
  TITLE     The complete genome sequence of the gram-positive bacterium
            Bacillus subtilis
  JOURNAL   Nature 390 (6657), 249-256 (1997)
  MEDLINE   98044033
REFERENCE   2  (bases 1 to 213420)
  AUTHORS   Kunst,F., Ogasawara,N., Yoshikawa,H. and Danchin,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur,
```

Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr, adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45 68 89 48

FEATURES Location/Qualifiers

source 1. .213420
/organism="Bacillus subtilis"
/strain="168"
/db_xref="taxon:1423"

gene complement(12. .1155)
/gene="sigA"

terminator complement(12. .36)
/gene="sigA"

CDS complement(40. .1155)
/gene="sigA"
/note="alternate gene name: rpoD, crsA"
/codon_start=1
/transl_table=11
/product="RNA polymerase major sigma-43 factor (sigma-A)"
/protein_id="CAB14463.1"
/db_xref="PID:e1183751"
/db_xref="PID:g2634967"
/db_xref="GI:2634967"
/db_xref="SWISS-PROT:P06224"
/translation="MADKQTHETELTFDQVKEQLTESGKKRGVLTYEEIAERMSSFEI
ESDQMDEYYEFLGEQGVELISENEETEDPNIQQLAKAEEEFDLNDLSVPPGVKINDPV
RMYLKEIGRVNLLSAKEEIAYAQKIEEGDEESKRRLAEANLRLVVSIAKRYVGRGMLF
LDLIHEGNMGLMKAVEKFDYRKGYKFSTYATWWIRQAITRAIADQARTIRIPVHMVET
INKLIRVQRQLLQDLGREPTPEEIAEDMDLTPEKVREILKIAQEPVSLETPIGEEDDS
HLGDFIEDQEATSPSDHAAYELLKEQLEDVLDTLTDREENVLRLRFGLDDGRTRTLEE
VGKVFGVTRERIRQIEAKALRKLRHPSRSKRLKDFLE"

gene complement(1354. .3165)
/gene="dnaG"

CDS complement(1354. .3165)
/gene="dnaG"
/EC_number="2.7.7.-"
/function="initiation of Okazaki fragments (DNA synthesis)"
/note="alternate gene name: dnaE"
/codon_start=1
/transl_table=11
/product="DNA primase"
/protein_id="CAB14464.1"
/db_xref="PID:e1183752"
/db_xref="PID:g2634968"
/db_xref="GI:2634968"
/db_xref="SWISS-PROT:P05096"
/translation="MGNRIPDEIVDQVQKSADIVEVIGDYVQLKKQGRNYFGLCPFHG
ESTPSFSVSPDKQIFHCFGCGAGGNVFSFLRQMEGYSFAESVSHLADKYQIDFPDDIT
VHSGARPESSGEQKMAEAHELLKKFYHHLLINTKEGQEALDYLLSRGFTKELINEFQI
GYALDSWDFITKFLVKRGFSEAQMEKAGLLIRREDGSGYFDRFRNRVMFPIHDHHGAV
VAFSGRALGSQQPKYMNSPETPLFHKSKLLYNFYKARLHIRKQERAVLFEGFADVYTA
VSSDVKESIATMGTSLTDDHVKILRRNVEEIILCYDSDKAGYEATLKASELLQKKGCK
VRVAMIPDGLDPDDYIKKFGGEKFKNDIIDASVTVMAFKMQYFRKGKNLSDEGDRLAY
IKDVLKEISTLSGSLEQEVYVKQLASEFSLSQESLTEQLSVFSKQNKPADNSGETKTR
RAHLTTKARQKRLRPAYENAERLLLAHMLRDRSVIKKVIDRVGFQFNIDEHRALAAYL
YAFYEEGAELTPQHLMARVTDDHISQLLSDILMLQVNQELSEAELSDYVKKVLNQRNW
SMIKEKEAERAEAERQKDFLRAASLAQEIVTLNRSLK"

gene complement(3199. .3789)
/gene="yqxD"

CDS complement(3199. .3789)
/gene="yqxD"
/function="unknown"
/note="alternate gene name: yqfM; similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB14465.1"
/db_xref="PID:e1183753"
/db_xref="PID:g2634969"
/db_xref="GI:2634969"
/db_xref="SWISS-PROT:P17868"
/translation="MSRISPGNFSTNSYTSTIIKDVRFCLLLCRSLMEGWRITLLNEQ
EKTIFVDADACPVKDEILQTASEYEVQVLFVASFEHYQLSRSNEEKWKYVDPHKEAAD
LYIANHVKPGDIVVTQDIGLASLLLNRNVSVMSERGRLYKEDTIDFALEGRHFSGKQR
RKGVYAKGPKKLNKEDRERFITLLQKILSNDEGILH"

gene complement(3947. .4759)
/gene="yqfL"

CDS complement(3947. .4759)
/gene="yqfL"
/function="unknown"
/note="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB14466.1"
/db_xref="PID:e1183754"
/db_xref="PID:g2634970"
/db_xref="GI:2634970"
/translation="MNNRIIYVVSDSVGETAELVVKAALSQFNGSADDIHVRRIPYVE
DIGTINEVISLAKADGGIICFTLVVPEIREYLIAEAEKANVLYYDIIGPLIDKMETAY
GLTAKYEPGRVRQLDEDYFKKVEAIEFAVKYDDGRDPRGILKADIVLIGVSRTSKTPL
SQYLAHKRLKVANVPIVPEVDPPEELFNVDPKKCIGLKISPDKLNHIRKERLKSLGLN
DKAIYANINRIKEELEYFEKIVDRIGCQVVDVSNKAVEETANIIHHLKTKNI"

gene complement(4785. .5423)
/gene="yqzB"

CDS complement(4785. .5423)
/gene="yqzB"
/function="unknown"
/note="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB14467.1"
/db_xref="PID:e1183755"
/db_xref="PID:g2634971"
/db_xref="GI:2634971"
/db_xref="SPTREMBL:O34994"
/translation="MSTIELNKRQEHILQIVKENGPITGEHIAEKLNLTRATILRPDLA
ILTMSGFLEARPRVGYFYTGKTGTQLLADKLKKLQVKDFQSIPVVIHENVSVYDAICT
MFLEDVGTLFVVDRDAVLVGVLSRKDLLRASIGQQELTSVPVHIIMTRMPNITVCRRE
DYVMDIAKHLIEKQIDALPVIKDTDKGFEVIGRVTKTNMTKILVSLSENEIL"

gene complement(5556. .7595)
/gene="glyS"

CDS complement(5556. .7595)
/gene="glyS"
/EC_number="6.1.1.14"
/note="alternate gene name: yqfK"
/codon_start=1
/transl_table=11
/product="glycyl-tRNA synthetase (beta subunit)"
/protein_id="CAB14468.1"
/db_xref="PID:e1183756"
/db_xref="PID:g2634972"
/db_xref="GI:2634972"
/db_xref="SWISS-PROT:P54381"
/translation="MSKQDLLLEIGLEEMPARFLNESMVQLGDKLTGWLKEKNITHGE
VKLFNTPRRLAVFVKDVAEKQDDIKEEAKGPAKKIALDADGNWTKAAIGFSKGQGANV
EDLYIKEVKGIEYVFVQKFQAGQETKSLLPELSGLITSLHFPKNMRWGNEDLRYIRPI
KWIVALFGQDVIPFSITNVESGRTTQGHRFLGHEVSIESPSAYEEQLKGQHVIADPSV
RKQMIQSQLETMAAENNWSIPVDEDLLDEVNHLVEYPTALYGSFESEFLSIPEEVLVT
TMKEHQRYFPVKDKNGDLLPHFITVRNGNSHAIENVARGNEKVLRARLSDASFFYKED
QKLNIDANVKKLENIVFHEELGSLADKVRRVTSIAEKLAVRLQADEDTLKHVKRAAEI
SKFDLVTHMIYEFPELQGIMGEKYARMLGEDEAVAAAVNEHYMPRSAGGETPSTFTGA
VVAMADKLDTIASFFSIGVIPTGSQDPYGLPRQASGIVAILLDRNWGISFEELLTFVQ
TDKENELLDFFTQRLKYVLNAEQIRHDVIDAVLESSELEPYSALHKAQVLEQKLGAPG
FKETAEALGRVISISKKGVRGDIQPDLFENEYEAKLFDAYQTAKENLQENFSKKDYEA
ALASLAALKEPIDAYFDHTMVIADNESLKANRLAQMVSLADEIKSFANMNALIVK"

gene complement(7588. .8475)
/gene="glyQ"

CDS complement(7588. .8475)
/gene="glyQ"
/EC_number="6.1.1.14"
/note="alternate gene name: yqfJ"
/codon_start=1
/transl_table=11
/product="glycyl-tRNA synthetase (alpha subunit)"
/protein_id="CAB14469.1"

```
                     /db_xref="PID:e1183757"
                     /db_xref="PID:g2634973"
                     /db_xref="GI:2634973"
                     /db_xref="SWISS-PROT:P54380"

  Query Match              2.7%;  Score 41.6;  DB 1;  Length 213420;
  Best Local Similarity   46.3%;  Pred. No. 0.76;
  Matches  137;  Conservative    0;  Mismatches  159;  Indels    0;  Gaps    0;

Qy     850 CTGAAGATGCTGCGCTGAGAGCAAAAACCTTTGCCATGGCAACAGAGGCAACATCAACCT 909
           ||||| |||      | | | | |  | || | | || | |  ||| | |||| |||
Db  164482 CTGAATATGTCAATATTACAACGATTATCTCTCCGATTGTAGAAGATGAAACAACAAAAA 164423

Qy     910 GCGAGGACCGGGTCACACATGCCCTGCACCAGATGAATAACGTACAACTGGTACATAATG 969
             || ||     || ||  |  | |  | ||  ||| | | | | || ||  |     |
Db  164422 TTGACGAAGACCTCTCATTTAACTTATATCATTTGATTGAGGAAAAAGTGAAAACGCTTA 164363

Qy     970 CAGAAAAAGGGGAATACGACAACAATCTCCAGGGGCTGGTTTCCACGGGGCGTGAGATGT 1029
             || ||||   || | | ||  || |  |    | |  || |      |  ||
Db  164362 TGAAAGAAGGTTTATCCAAAAAGGATATTAACCAGTACATCTTGACTGATGTCCATTTGC 164303

Qy    1030 TCCGCCTGGCAACACTGGAACAGATTGCCCGGGAAAAAGCCGGAACACTGGCTTTAGTCG 1089
              |  | ||     ||     |||||| |      ||| | || || | |
Db  164302 ATGTCAGAAGTTTTTTTCATCATCAAGCTTTTCAAAAAGACAACCTGCTGACATTCGTTG 164243

Qy    1090 ATGACGTTGAGGTCTATCTGGCGTTCCAGAATAAGCTGAAGGAATCACTTGAGCTG 1145
           | |||| || | |  |  || |    |||   ||     |  |    || |||||||
Db  164242 AAGACGATGTGATTCAAATGACAAAACAGTTAAAAGAAATAGCTGAACATGAGCTG 164187


RESULT  13
GORINVOLUA/c
LOCUS       GORINVOLUA   1908 bp    DNA             PRI       07-SEP-1994
DEFINITION  Gorilla gorilla involucrin gene large allele, complete cds.
ACCESSION   M23603 J04499
NID         g340977
VERSION     M23603.1  GI:340977
KEYWORDS    epidermal protein; involucrin.
SOURCE      Gorilla gorilla (individual_isolate Gorilla F) DNA.
  ORGANISM  Gorilla gorilla
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Gorilla.
REFERENCE   1  (bases 1 to 1908)
  AUTHORS   Teumer,J. and Green,H.
  TITLE     Divergent evolution of part of the involucrin gene in the
            hominoids: Unique intragenic duplications in the gorilla and human
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 86, 1283-1286 (1989)
  MEDLINE   89145226
COMMENT     Draft entry and computer readable copy of sequence [1] kindly
            provided by J.K. Teumer (03-31-89).
FEATURES             Location/Qualifiers
     source          1. .1908
                     /organism="Gorilla gorilla"
                     /isolate="Gorilla F"
                     /db_xref="taxon:9593"
                     /cell_type="fibroblast"
                     /clone="pIGorH6.2"
                     /note="(vector pGEM3)"
     CDS             1. .1908
                     /codon_start=1
                     /product="involucrin"
                     /protein_id="AAA35471.1"
                     /db_xref="PID:g536827"
                     /db_xref="GI:536827"
                     /translation="MSLQHTLPVTLSPALSQELLKTVPPPVNTQQEQMKPPTPLPPPC
                     QKVPVKLPVEVPSKQEEKHMTAVKGLPEQECEQQQQEPQEQEPQQQHWEQHEEHQKAE
                     NPEQQLKQEKAQRDQQLNKQLEEEKKLLDQQLDQELVKRDEQLGMKKEQLLELPEQRE
                     GELKHLEQQKGQLELPEQQEGQLELPEQQEGQLKHLEQQKGQLELPEQQEGQLELPEQ
                     QEGQLELPEQQEGQLKHLEQQEGQLELPEQQEGQLKHLEQQKGQLELPEQQEGQLELP
                     EQQEGQLELPEQQEGQLKHLEQQEGQLKHLENQEGQLEVPEEQVGQLNYLEQQEGQLK
                     HLDQQEKQPELPEQQVGQLKHLEKQEGQPKHLEQQEGQLEHLEEQEGQLKHLEQQEGQ
                     LEHLEHQEGQLGLPEQQVRQLKQLEKEEGQPKHLEEEEGQLKHLVQQEGQLEHLVQQE
                     GQLEHLVQQEGQLEQQEGQVEHLEQQVGQLKHLEEQEGQLKHLEQQQGQLGVPEQQVG
                     QPNNLEQEEKQLELPEQQEGQLKHLEKQEAQLELPEQQVGQPKHLEQQEKQLEHPEQQ
                     DGQLKHLEQQEGQLKDLEQQKGQLEQPVFAPAPGQVQDIQPALPTKGELLLPVEHQQQ
                     KQEVQWPPKHK"
BASE COUNT      554 a    476 c    696 g    182 t
ORIGIN


  Query Match              2.6%;  Score 40.6;  DB 9;  Length 1908;
  Best Local Similarity   46.1%;  Pred. No. 0.99;
  Matches  171;  Conservative    0;  Mismatches  199;  Indels    1;  Gaps    1;

Qy     126 CCTGCCGGAGTTGCCGGAAGGACTACGGGAGCTGGAGGTCTCTGGTAACCTACAACTGAC 185
           ||| | |  ||| | || |   |  | |   |    | ||    |  |||| || ||| |
Db     991 CCTCCTGCTGTTCCAGGTAATTCAGCTGCCCCACCTGCTCCTCTGGGACCTCCAGCTGCC 932

Qy     186 CAGCCTGCCATCGCTGCCGCAGGGACTACAGAAGCTGTGGGCCTATAATAATTGGCTGGC 245
           |  ||||      | |  ||     || |     ||| |  | |       |  |||| |
Db     931 CCTCCTGGTTCTCCAGATGCTTCAGCTGCCCCTCCTGCTGCTCCAGGTGCTTCAGCTGTC 872

Qy     246 CAGCCTGCCGACGTTGCCGCCAGGACTAGGGGATCTGGCGGTCTCTAATAACCAGCTGAC 305
           |  ||||| |   | |  ||     ||       |||  | |||   |   |||||||  |
Db     871 CCTCCTGCTGCTCTGGGAGCTCCAGCTGCCCCTCCTGCTGCTCTGGGAGCTCCAGCTGCC 812

Qy     306 CAGCCTGCCGGAGATGCCGCCAGCACTACGGGAGCTGAGGGTCTCTGGTAACAACCTGAC 365
           |  ||||| |      |  ||     || |     |||  | ||   |       | ||| |
Db     811 CCTCCTGCTGCTCTGGGAGCTCCAGCTGCCCCTTCTGCTGCTCCAGGTGCTTCAGCTGTC 752

Qy     366 CAGCTGCGCGCGCTGCCGTCAGGA-CTACAGAAGCTGTGGGCCTATAATAATCGGCTGAC 424
           |  |    || |||     |   | || |     |||  | |||      || |||| |
Db     751 CCTCCTGCTGCTCTGGGAGCTCCAGCTGCCCCTCCTGCTGCTCTAGGTGCTTCAGCTGTC 692

Qy     425 CAGCCTGCCGGAGATGTCGCCAGGACTACAGGAGCTGGATGTCTCTCATAACCAGCTGAC 484
           |  ||||| |      |  ||     || |     |||    |||   |  |||||||  |
Db     691 CCTCCTGCTGCTCTGGGAGCTCCAGCTGCCCCTCCTGCTGCTCTGGGAGCTCCAGCTGCC 632

Qy     485 CCGCCTGCCGC 495
           || ||||| ||
Db     631 CCTCCTGCTGC 621


RESULT  14
SC10A7
LOCUS       SC10A7      39739 bp    DNA             BCT       09-JUN-1999
DEFINITION  Streptomyces coelicolor cosmid 10A7.
ACCESSION   AL078618
NID         g5042256
VERSION     AL078618.1  GI:5042256
KEYWORDS    acetyltransferase; bi-functional transferase/deacetylase; electron
            transfer oxidoreductase; fumarylacetoacetase; helicase; hydrolase;
            integral membrane protein; NADH dehydrogenase subunit; NLP/P60
            family secreted protein; nuoA; nuoB; nuoC; nuoD; nuoE; nuoF; nuoG;
            nuoH; nuoI; nuoJ; nuoK; nuoL; nuoM; nuoN; polypetide deformylase;
            type IV peptidase; ubiquinone/menaquinone methyltransferase.
SOURCE      Streptomyces coelicolor.
  ORGANISM  Streptomyces coelicolor
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 39739)
  AUTHORS   Redenbach,M., Kieser,H.M., Denapaite,D., Eichner,A., Cullum,J.,
            Kinashi,H. and Hopwood,D.A.
  TITLE     A set of ordered cosmids and a detailed genetic and physical map
            for the 8 Mb Streptomyces coelicolor A3(2) chromosome
  JOURNAL   Mol. Microbiol. 21 (1), 77-96 (1996)
  MEDLINE   97000351
REFERENCE   2  (bases 1 to 39739)
  AUTHORS   Saunders,D.C. and Harris,D.
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 39739)
  AUTHORS   Bentley,S.D., Parkhill,J., Barrell,B.G. and Rajandream,M.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUN-1999) Streptomyces coelicolor sequencing project,
            Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
```

David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK

COMMENT Notes:
Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics
Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.
(URL; http://www.sanger.ac.uk/Projects/S_coelicolor/)
CDS are numbered using the following system eg SC7B7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).
The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.
The length in codons is given for each CDS.
Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database.
The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon.
Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at http://www.nih.go.jp/jun/cgi-bin/frameplot.pl. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or (att)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.
IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.
Cosmid 10A7 lies to the left of D13 on the AseI-D genomic restriction fragment.

FEATURES Location/Qualifiers

source 1. .39739
/organism="Streptomyces coelicolor"
/strain="A3(2)"
/db_xref="taxon:1902"
/clone="cosmid 10A7"

gene 565. .1056
/gene="SC10A7.01"

CDS 565. .1056
/gene="SC10A7.01"
/note="SC10A7.01, unknown, len: 163aa; some similarity to proteins associated with bacteriophages eg. TR:Q38352 (EMBL:L29568) ORF360 hypothetical protein from Lactobacillus delbrueckii subsp. lactis bacteriophage LL-H (360 aa) fasta scores; opt: 255, z-score: 294.2, E(): 4.7e-09, (45.5% identity in 101 aa overlap)"
/codon_start=1
/transl_table=11
/label=SC10A7.01
/product="hypothetical protein"
/protein_id="CAB44511.1"
/db_xref="PID:e1487133"
/db_xref="PID:g5042257"
/db_xref="GI:5042257"
/translation="MSANGQNRHARTSRLARKLAVAGTGAAVLALPLIGATTASAATP
AAATTATTAATAYPDNLDGWIRESLDIMAQHGIPGTYEGIHRNIMRESSGNPLAINNW
DINAINGTPSKGLLQVIDPTFQAYHVEGTSWDPYDPVANITAACNYAADRYGSMDNVF
SAY"

gene complement(1080. .1889)
/gene="SC10A7.02c"

CDS complement(1080. .1889)
/gene="SC10A7.02c"
/note="SC10A7.02c, conserved hypothetical protein, len: 269aa; similar to many hypothetical proteins egs. TR:CAA22232 (EMBL:AL034355) (280 aa) fasta scores; opt: 1028, z-score: 1143.6, E(): 0, (61.5% identity in 260 aa overlap) and TR:CAA22222 (EMBL:AL034355) (305 aa) fasta scores; opt: 546, z-score: 611.4, E(): 1e-26, (40.2% identity in 271 aa overlap) hypothetical proteins from Streptomyces coelicolor. Also similar to SW:YIDA_ECOLI hypothetical protein from Escherichia coli (270 aa) fasta scores; opt: 314, z-score: 356.2, E(): 1.6e-12, (31.0% identity in 268 aa overlap). Contains Pfam match to entry PF00592 DUF3, Domain of Unknown Function 3, score 116.90, E-value 3.9e-31."
/codon_start=1
/transl_table=11
/label=SC10A7.02c
/product="conserved hypothetical protein"
/protein_id="CAB44512.1"
/db_xref="PID:e1487134"
/db_xref="PID:g5042258"
/db_xref="GI:5042258"
/translation="MDRTAYSLVATDLDGTLLRGDDTVSDRSLAALARVAGAGARHLV
VTGRPAPRVRPLLDRLGCTGLAVCGQGAQVYDAGAHRMLWSVTLDRELAETALGKIEA
EVGQVHAAVDQDGVDGLTLIEPDYLMPHPTLPAVRVERRAQLWSTPISKVLLRHPELT
DDELAATARAVVGSLATVTMSGPGTVELQPCGITKATGLALAAEHLGLERRRTIAFGD
MPNDIPMFQWAAHGVAMAGAHPELKAVADEVTTTNEDDGVAVVLERIFGTS"

misc_feature complement(1125. .1868)
/gene="SC10A7.02c"
/note="Pfam match to entry PF00592 DUF3, Domain of Unknown Function 3, score 116.90, E-value 3.9e-31."

gene 2374. .3630
/gene="SC10A7.03"

CDS 2374. .3630
/gene="SC10A7.03"
/note="SC10A7.03, probable fumarylacetoacetase, len: 418aa; similar to several from eukaryotes eg. SW:FAAA_HUMAN fumarylacetoacetase from human (419 aa) fasta scores; opt: 1192, z-score: 1370.0, E(): 0, (44.6% identity in 426 aa overlap)."
/codon_start=1
/transl_table=11
/label=SC10A7.03
/product="putative fumarylacetoacetase"
/protein_id="CAB44513.1"
/db_xref="PID:e1487135"
/db_xref="PID:g5042259"
/db_xref="GI:5042259"
/translation="MPPFDVPEGDPFGPHNLPYGVFSIPGSLDRTDRTDRTVGVRLGD
HVLDAGAAAHALGSPYAALLAQPTLNPLLAAGRTAWSDVRRALTAWVTVPSHRETLEP
FFHPLSSVTLHLPFEVADYVDFYASENHARNVGQIFRPDAADSLTPNWKHLPIGYHGR
SGTVVVSGTDVVRPSGQRKAPADAAPVFGPSVRLDIEAEVGFVVGVPSELGSPVALGD
FREHVFGLCLLNDWSARDVQAWEYVPLGPFLGKSFTTSVSAWITPLDALEEARVAPPE
RTHELLPYLDDTDTDQDAPGGYDLRISVAVNGHVVSEPPFSTMYWTAAQQLAHMTVNG
ASLRTGDLYGSGTVSGPTERERGSLLELTWNGRDPLELPNGKRTFLEDGDVVTLSAWA
PGPDGVRVGLGEVTGRVVGRGVEQGV"

gene 3937. .4875
/gene="SC10A7.04"

CDS 3937. .4875
/gene="SC10A7.04"
/note="SC10A7.04, possible integral membrane protein, len: 312aa; similar to others egs. TR:O32910 (EMBL:AL008609) hypothetical protein from Mycobacterium leprae (316 aa) fasta scores; opt: 405, z-score: 470.7, E(): 6.9e-19, (32.7% identity in 324 aa overlap) and TR:O86628 (EMBL:AL031155) putative integral membrane protein from Streptomyces coelicolor (309 aa) fasta scores; opt: 222, z-score: 261.4, E(): 3.1e-07, (31.0% identity in 310 aa overlap). Contains possible membrane spanning hydrophobic regions."
/codon_start=1
/transl_table=11
/label=SC10A7.04
/product="putative integral membrane protein"
/protein_id="CAB44514.1"
/db_xref="PID:e1487136"
/db_xref="PID:g5042260"
/db_xref="GI:5042260"
/translation="MTVCLLLLFVVAVTAAVPVPRALTRSAWPERDPVVGLWVWQCLV
ATVLLCCLTALILGATAVFGTVRAQLFAPAPPAVTAAYNLSAGPVWAAALTLLLAAGA

```
                     AWTTAMLARELVEARRRRARSRAHLRERAPDLPAGLPAARGPLLVLEDEYPDAWWMPG
                     SPPQLIVTTGALHRLTDHQLDAVLTHERGHARARHDWLLHLSTALATGFPSVPLFAHF
                     CDQTHRLVELSADDTASRRCGHLTTALALIELNQHRGVLSCASSHRLLGERVDRLLEP
                     RPRLPRRHRALTTATAALVPLIPLLITFAPGLTALA"
     gene            4937. .5161
                     /gene="SC10A7.05"
     CDS             4937. .5161
                     /gene="SC10A7.05"
                     /note="SC10A7.05, unknown, len: 74aa; predicted based upon
                     amino acid and codon usage."
                     /codon_start=1
                     /transl_table=11
                     /label=SC10A7.05
                     /product="hypothetical protein"
                     /protein_id="CAB44515.1"
                     /db_xref="PID:e1487137"
                     /db_xref="PID:g5042261"
                     /db_xref="GI:5042261"
                     /translation="MISTPRCASRRCAGPDRAVPVLAKARADRNADVRKSAVLAPTRH
                     TDAAEDARTALATATGDTDAAGRAYATRAL"
     gene            complement(5200. .7230)
                     /gene="SC10A7.06c"
     CDS             complement(5200. .7230)

  Query Match              2.6%;  Score 40;  DB 1;  Length 39739;
  Best Local Similarity   53.1%;  Pred. No. 1.8;
  Matches   85;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy     624 GGGGCCTTCCGTCCCCGGGAAGCCCGGGCACTGCACCTGGCGGTCGCTGACTGGCTGACG 683
           || |||    |  ||| |    || |||| |  |||| | ||| ||| |   | |
Db   17537 GGCGCCGCGCTCCCCGGCCCGGCGCGGGATCCACACCAGCCGGGCGCCGGTCGCCGCCGA 17596

Qy     684 TCTGCACGGGAGGGGGAAGCGGCCCAGGCAGACAGATGGCAGGCGTTCGGACTGGAAGAT 743
           | |    |  || |  | ||||   |   || || | ||| |||||      || |||
Db   17597 GGTCCGTACGGCGGAGGTGAGGCCACCGGCCACGGACGCCAGCCGTTCACCGACGACGAT 17656

Qy     744 AACGCCGCCGCCTTCAGCCTGGTCCTGGACAGACTGCGTG 783
           ||| ||||  |  |   | || ||| | | | ||| |||
Db   17657 CACGGCGCCCTCGGCGCGCAGGGCCTCGGCGGCCTGGGTG 17696


RESULT  15
YEPYOPM
LOCUS       YEPYOPM      1397 bp    DNA             BCT       21-JUN-1994
DEFINITION  Y.pestis outer membrane protein (yopM) gene, complete cds.
ACCESSION   M25810
NID         g155539
VERSION     M25810.1  GI:155539
KEYWORDS    outer membrane protein.
SOURCE      Y.pestis DNA.
  ORGANISM  Yersinia pestis
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Yersinia.
REFERENCE   1  (bases 1 to 1397)
  AUTHORS   Leung,K.Y. and Straley,S.C.
  TITLE     The yopM gene of Yersinia pestis encodes a released protein having
            homology with the human platelet surface protein GPIb-alpha
  JOURNAL   J. Bacteriol. 171, 4623-4632 (1989)
  MEDLINE   89359090
COMMENT     Draft entry and computer-readable sequence for [1] kindly submitted
            by S.C.Straley, 06-JUL-1989.
FEATURES             Location/Qualifiers
     source          1. .1397
                     /organism="Yersinia pestis"
                     /db_xref="taxon:632"
     CDS             164. .1267
                     /note="outer membrane protein (yopM)"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA27670.1"
                     /db_xref="PID:g155540"
                     /db_xref="GI:155540"
                     /translation="MFINPRNVSNTFLQEPLRHSSNLTEMPVEAENVKSKTEYYNAWS
                     EWERNAPPGNGEQREMAVSRLRDCLDRQAHELELNNLGLSSLPELPPHLESLVASCNS
                     LTELPELPQSLKSLLVDNNNLKALSDLPPLLEYLGVSNNQLEELPELQNLPFLTAIYA
                     DNNSLKKLPDLPLSLESIVAGNNILEELPELQNLPFLTTIYADNNLLKTLPDLPPSLE
                     ALNVRDNYLTDLPELPQSLTFLDVSENIFSGLSELPPNLYYLNASSNEIRSLCDLPPS
                     LEELNVSNNKLIELPALPPRLERLIASFNHLAEVPELPQNLKQLHVEYNPLREFPDIP
                     ESVEDLRMNSERVVDPYEFAHETTDKLEDDVFE"
BASE COUNT      450 a    274 c    257 g    416 t
ORIGIN


  Query Match              2.6%;  Score 39.8;  DB 1;  Length 1397;
  Best Local Similarity   48.1%;  Pred. No. 1.6;
  Matches  142;  Conservative    0;  Mismatches  152;  Indels    1;  Gaps    1;

Qy     190 CTGCCATCGCTGCCGCAGGGACTACAGAAGCTGTGGGCCTATAATAATTGGCTGGCCAGC 249
           ||||||    | |||||| |  ||       |    |  | | | |||    |  |  |
Db     851 CTGCCAGAATTACCGCAGAGTTTAACCTTCTTAGATGTTTCTGAAAATATTTTTTCTGGA 910

Qy     250 CTGCCGACGTTGCCGCCAGGACTAGGGGATCTGGCGGTCTCTAATAACCAGCTGACCAGC 309
           |  ||   ||||| |||    |      ||||    |  || |  ||  |  | |    |
Db     911 TTATCGGAATTGCCACCAAACTTGTATTATCTCAATGCATCCAGCAATGAAATAAGATCC 970

Qy     310 CTGCCGGAGATGCCGCCAGCACTACGGGAGCTGAGGGTCTCTGGTAACAACCTGA-CCAG 368
           |    ||  | || ||  |||| |  ||| |  ||| ||  ||| | | |
Db     971 TTATGCGATTTACCCCCTTCACTGGAAGAACTTAATGTCAGTAATAATAAGTTGATCGAA 1030

Qy     369 CTGCGCGCGCTGCCGTCAGGACTACAGAAGCTGTGGGCCTATAATAATCGGCTGACCAGC 428
           ||||  ||| | ||  || |  || |      |    || |    |||||  ||  |
Db    1031 CTGCCAGCGTTACCTCCACGCTTAGAACGTTTAATCGCTTCATTTAATCATCTTGCTGAA 1090

Qy     429 CTGCCGGAGATGTCGCCAGGACTACAGGAGCTGGATGTCTCTCATAACCAGCTGA 483
           | || ||  || ||| |   ||  |  ||||  | ||     | ||||  ||||
Db    1091 GTACCTGAATTGCCGCAAAACCTGAAACAGCTCCACGTAGAGTACAACCCTCTGA 1145


Search completed: October  5, 1999, 18:05:07
Job time: 11750 sec
```

```
OM of: US-09-068-804-12 to: EST:*   out_format : pfs

Date: Oct 5, 1999  7:34 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn2_1/USPTO_spool/US09068804/runat_05101999_145134_10306/app_query.fasta.1
 -DB=EST -QFMT=fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=pam150 -TRANS=human40.cdi
 -LIST=45 -DOCALIGN=200 -THR_SCORE=escore -ALIGN=15 -MODE=LOCAL
 -OUTFMT=pfs -NORM=stat -USER=US09068804 -NCPU=6 -ICPU=3 -WAIT
 -THREADS=1

Search information block:
Query: US-09-068-804-12
Query length: 239
Database: EST:*
Database sequences: 2546578
Database length: 986266752
Search time (sec): 940.790000


score_list:
Sequence              Strd Orig    ZScore    EScore Len  ! Documentation  ..
gb_est14:AA410533       +   83.00  189.53  0.0483    465  ! AA410533 zv23c04.r1 Soares_NhHM
gb_est16:AA548563       -   82.50  186.27  0.0734    557  ! AA548563 nk50c06.s1 NCI_CGAP_GC
em_est21:HSM005936      +   82.00  184.17  0.0961    600  ! Al041460 Homo sapiens mRNA; EST
gb_est30:AI638108       -   76.50  172.86  0.4101    466  ! AI638108 tt24d10.x1 NCI_CGAP_GC
gb_est11:AA232769       +   75.50  170.30  0.5696    466  ! AA232769 zr45f04.r1 Soares_NhHM
gb_est26:AI324939       -   76.00  168.91  0.6803    594  ! AI324939 mb50h06.x1 Soares mous
gb_est27:AI431844       -   75.50  168.70  0.6992    539  ! AI431844 ti25h05.x1 NCI_CGAP_Ki
gb_est27:AI442505       -   73.50  167.56  0.8091    375  ! AI442505 sa27h06.y1 Gm-c1004 Gl
gb_est12:AA314946       +   74.00  165.70  1.03      499  ! AA314946 EST186775 HCC cell lin
gb_est12:AA296865       +   72.50  165.57  1.04      356  ! AA296865 EST112446 Aorta endoth
gb_est17:C70185         +   72.50  165.45  1.06      360  ! C70185 C70185 Yuji Kohara unpub
gb_est17:AA627031       +   72.50  165.15  1.10      370  ! AA627031 MBAFCW2G01T3 Brugia ma
gb_est19:AA805956       +   72.00  164.99  1.13      334  ! AA805956 oc20b03.s1 NCI_CGAP_GC
gb_est11:AA257490       +   73.50  164.85  1.15      480  ! AA257490 MBAFCZ0E10T3 Brugia ma
gb_est22:AA995560       -   73.50  164.67  1.17      488  ! AA995560 os02a12.s1 NCI_CGAP_Lu
gb_est26:AI396994       +   73.50  164.56  1.19      493  ! AI396994 fb08b12.y1 zebrafish f
gb_est19:AA793297       +   71.50  164.04  1.27      324  ! AA793297 vm13g07.r1 Knowles Sol
gb_est24:AI222037       +   72.00  163.78  1.31      373  ! AI222037 qh01h02.x1 Soares_NFL_
gb_est21:AA967200       +   72.50  162.73  1.50      461  ! AA967200 ua50f05.r1 Stratagene
gb_est9:AA123608        +   72.00  162.58  1.53      416  ! AA123608 MBAFCE9C04T3 Brugia ma
gb_est10:AA183907       +   72.00  162.58  1.53      416  ! AA183907 mt21h01.r1 Soares mous
gb_est20:AB011386       +   72.00  161.79  1.70      447  ! AB011386 AB011386 Rainbow trout
gb_est32:AI739126       -   73.00  161.01  1.88      606  ! AI739126 wi18d01.x1 NCI_CGAP_Co
gb_est15:AA474977       +   72.50  160.57  1.98      561  ! AA474977 vh08e12.r1 Soares mous
gb_est23:AI154232       +   72.00  160.54  1.99      501  ! AI154232 ud30e05.r1 Soares mous
gb_est8:AA009519        +   72.00  160.34  2.04      510  ! AA009519 ze82b09.r1 Soares_feta
gb_est7:W77396          +   72.00  159.96  2.14      528  ! W77396 me62e05.r1 Soares mouse
gb_est9:AA109308        +   71.00  159.58  2.25      433  ! AA109308 MBAFCE0A01T3 Brugia ma
gb_est7:W58964          +   69.00  159.24  2.35      280  ! W58964 md58h11.r1 Soares mouse
gb_est13:AA355551       +   69.50  159.23  2.36      315  ! AA355551 EST63970 Jurkat T-cell
gb_est21:AA984356       -   70.50  158.95  2.44      408  ! AA984356 am84f03.s1 Stratagene
gb_est15:AA510251       +   72.00  158.83  2.48      585  ! AA510251 vh58a07.r1 Soares mous
gb_est32:AI723694       +   72.00  158.19  2.69      620  ! AI723694 BSBmL3SZ45G8SK Brugia
gb_est26:AI324391       +   72.50  157.65  2.88      732  ! AI324391 mb50h06.y1 Soares mous
gb_est29:AI573408       -   72.00  157.63  2.89      653  ! AI573408 mn97h09.x1 Stratagene
gb_est12:AA285586       +   71.50  157.50  2.94      588  ! AA285586 vb83g11.r1 Soares mous
gb_est22:AI026924       -   70.50  156.94  3.16      490  ! AI026924 ov98b10.x1 Soares_test
gb_est31:AV029710       -   66.00  156.91  3.17      172  ! AV029710 AV029710 Mus musculus
gb_est14:AA388296       +   70.50  156.67  3.27      502  ! AA388296 vc24f02.r1 Ko mouse em
gb_est14:AA450502       +   70.50  156.18  3.48      525  ! AA450502 vf77h05.r1 Soares mous
gb_est16:AA543229       +   70.50  155.75  3.68      546  ! AA543229 vk37g01.r1 Soares mous
gb_est22:AI021580       -   69.50  155.29  3.91      451  ! AI021580 ub13c10.r1 Soares mous
gb_est21:AA976793       -   70.50  155.12  3.99      578  ! AA976793 oq20b11.s1 NCI_CGAP
gb_est2:R14169          +   69.50  154.93  4.09      466  ! R14169 yf78f01.r1 Soares inf
gb_est9:AA104666        -   70.50  154.90  4.11      590  ! AA104666 mo46h08.r1 Life Tec

seq_name: gb_est14:AA410533

seq_documentation_block:
LOCUS       AA410533      465 bp    mRNA            EST       18-MAY-1997
DEFINITION  zv23c04.r1 Soares_NhHMPu_S1 Homo sapiens cDNA clone IMAGE:754470 5'
            similar to gb:M25393 T-CELL PROTEIN-TYROSINE PHOSPHATASE (HUMAN);,
            mRNA sequence.
ACCESSION   AA410533
NID         g2069657
VERSION     AA410533.1  GI:2069657
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 465)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J.,
            Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-Merck EST Project 1997
  JOURNAL   Unpublished (1997)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1404744.

            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 451.
FEATURES             Location/Qualifiers
     source          1. .465
                     /organism="Homo sapiens"
                     /db_xref="GDB:5977373"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:754470"
                     /clone_lib="Soares_NhHMPu_S1"
                     /tissue_type="Pooled human melanocyte, fetal heart, and
                     pregnant uterus"
                     /lab_host="DH10B"
                     /note="Organ: mixed (see below); Vector: pT7T3D-Pac
                     (Pharmacia) with a modified polylinker; Site_1: Not I;
                     Site_2: Eco RI; Equal amounts of plasmid DNA from three
                     normalized libraries (melanocyte 2NbHM, pregnant uterus
                     NbHPU, and fetal heart NbHH19W) were mixed, and ss circles
                     were made in vitro. Following HAP purification, this DNA
                     was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from pools of
                     5,000 clones made from the same 3 libraries. The pools
                     consisted of I.M.A.G.E. clones 260232-265223,
                     340488-345479, and 484488-489479."
BASE COUNT       146 a     89 c     99 g    131 t
ORIGIN


alignment_scores:
           Quality:   83.00              Length:    123
             Ratio:  1.064                Gaps:      9
 Percent Similarity:  63.415   Percent Identity: 30.894

alignment_block:
 US-09-068-804-12 x AA410533   ..

 Align seg 1/1  to: AA410533  from: 1  to: 465
```

```
    114 TyrAsnMetGlnLeu.................SerCysGlyGluLysAr 124
        |||:::::::::|||                 ::::::||||||:::||
     47 TATACAGTACATCTACTACAATTAGAAAATATCAATAGTGGTGAAACCAG 96

    124 gTyrThrIleProValLeuHisValLysAsnTrpProAspHis...GlnP 140
        |  ||||||:::   :::|||   ::::::|||||||||         |
     97 A...ACAATATCTCACTTTCATTATACTACCTGGCCAGATTTTGGAGTCC 143

    140 roLeuProSerThrAspGlnLeuGluTyrLeuAlaAspArgValLysAsn 156
        ||   ::::::::::::   |||:::::::|||      :::|||::::::
    144 CTGAATCACCAGCTTCATTTCTCAATTTCTTGTTT...AAAGTGAGAGAA 190

    157 SerAsnGlnAsnGlyAlaProGlyArgSerSerSerAspLysHisLeuPr 173
        |||                 |||      :::::::|||         ::
    191 TCT.................GGCTCCTTGAACCCTGACCATGGGCCTGC 222

    173 oMetIleHisCysLeuGlyGlyValGlyArgThrGlyThr.........M 187
        ::::|||||||||   :::|||:::|||||||:::||||||        :
    223 GGTGATCCACTGTAGTGCAGGCATTGGGCGCTCTGGCACCTTCTCTCTGG 272

    187 etAlaAlaAlaLeuValLeuLysAspAsnProHisSer...AsnLeuGlu 202
        ::::::::   |||||||||:::::::::   ::::::   |||:::
    273 TAGACACTTGTCTTGTTTTGATGGAAAAAGGAGATGATATTAACATAAAA 322

    203 GlnValArgAlaAspPheArgIleHis......GlyThrIleGluCysTr 217
        ||||||      :::   |||   :::      |||   |||:::
    323 CAAGTGTTACTGAACATGAGAAAATACCGAATGGGTCTTATTCAG..... 367

    217 pLysThrProAlaGlnPhe 223
          ||||||:::|||:::
    368 ....ACCCCAGATCAACTG 382
```

seq_name: gb_est16:AA548563

seq_documentation_block:

```
LOCUS       AA548563      557 bp    mRNA            EST       11-AUG-1997
DEFINITION  nk50c06.s1 NCI_CGAP_GC2 Homo sapiens cDNA clone IMAGE:1016938 3'
            similar to gb:M68941 PROTEIN-TYROSINE PHOSPHATASE PTP-MEG (HUMAN);,
            mRNA sequence.
ACCESSION   AA548563
NID         g2318845
VERSION     AA548563.1  GI:2318845
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 557)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1400780.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: Stratagene, Inc., David B. Krizman,
            Ph.D.
             cDNA Library Arraying: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 390.
FEATURES             Location/Qualifiers
     source          1. .557
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1016938"
                     /clone_lib="NCI_CGAP_GC2"
                     /tissue_type="germ cell tumor"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; Cloned unidirectionally.  Primer: Oligo dT. Bulk
                     germ cell tumor. 5' adaptor sequence: 5' GAATTCGGCACGAG 3'
                     3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'
                     Average insert size:  1.2 kb."
BASE COUNT      148 a    118 c    113 g    178 t
ORIGIN
```

```
alignment_scores:
           Quality:   82.50              Length:    159
             Ratio:  1.100                Gaps:      9
 Percent Similarity:  47.170   Percent Identity: 22.013

alignment_block:
 US-09-068-804-12 x AA548563/rev   ..

 Align seg 1/1  to reverse of: AA548563  from: 1  to: 557

    115 AsnMetGlnLeuSerCysGlyGluLysArgTyrThrIleProValLeuHi 131
        :::|||   |||         ||||||      :::   |||:::   ::
    466 AAGATGACCCTATTTAACCAAGAGAAAAATGAAAGTCGTCCACTCACTCA 417

    131 sValLys......AsnTrpProAspHis...GlnProLeuProSerThrA 145
        :::::::      :::|||||||||||||      |||      |||:::|
    416 GATCCAGTACATAGCCTGGCCTGACCATGGAGTCCCTGATGATTCGAGTG 367

    145 spGlnLeuGluTyrLeuAlaAspArgValLysAsnSerAsnGlnAsnGly 161
        ||   |||:::::::::      :::|||:::|||
    366 ACTTTCTAGATTTTGTTTGT...CATGTACGAAACAAG..... 332

    162 AlaProGlyArgSerSerSerAspLysHisLeuProMetIleHisCysLe 178
                |||::::::   :::      :::   ::::::|||||
    331 .........AGGGCTGGCAAGGAAGAACCCGTT...GTTGTCCATTGCAG 294

    178 uGlyGlyValGlyArgThrGly.........ThrMetAlaAlaAlaLeuV 192
         :::|||:::||||||||||||         ||||||::::::|||:::
    293 TGCTGGAATCGGAAGAACTGGGGTTCTTATTACTATGGAAACAGCCATGT 244

    192 alLeuLysAsp..................................... 195
          |||   :::
    243 GTCTCATTGAATGCAATCAGCCAGTTTATCCACTAGATATTGTAAGAACA 194

    196 ......AsnProHisSerAsnLeuGluGlnValArgAlaAspPheArg.. 209
              ::::::::::::   :::   |||:::   ::::::::|||
    193 ATGAGAGATCAGCGAGCCATGATGATCCAAACACCTAGTCAATACAGATT 144

    210 .IleHisGlyThrIle.............................. 214
         :::      :::|||
    143 TGTATGTGAAGCTATTTTGAAAGTTTATGAAGAAGGCTTTGTTAAACCCT 94

    215 .....................................GluCys 216
                                             |||
     93 TAACAACATCAACAAATAAATAAGAAAGCAAAAAGATCTGGGATATGTGT 44

    217 TrpLysThrProAlaGlnPheValGln 225
        |||||||||:::   ::::::|||::: 
     43 TGGAAAACTGCTTTACCTTATGTTCAC 17

seq_name: em_est21:HSM005936

seq_documentation_block:
ID   HSM005936  standard; RNA; EST; 600 BP.
XX
AC   AL041460;
XX
SV   AL041460.1
XX
NI   e1399797
```

```
XX
DT   12-MAR-1999 (Rel. 59, Created)
DT   12-MAR-1999 (Rel. 59, Last updated, Version 1)
XX
DE   Homo sapiens mRNA; EST DKFZp434E1817_r1  (from clone
DE   DKFZp434E1817)
XX
KW   EST; expressed sequence tag.
XX
OS   Homo sapiens (human)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria;
OC   Primates; Catarrhini; Hominidae; Homo.
XX
RN   [1]
RP   1-600
RA   Bloecker H., Boecher M., Brandt P., Mewes W., Gassenhuber J.,
RA   Wiemann S.;
RT   ;
RL   Submitted (12-MAR-1999) to the EMBL/GenBank/DDBJ databases.
RL   MIPS, Am Klopferspitz 18a D-82152 Martinsried, GERMANY
XX
CC   Clone from S. Wiemann, sequenced by GBF within the cDNA
CC   sequencing consortium of the German Genome Project
CC   s1 sequence also available
CC   This clone is available at the RZPD in Berlin
CC   Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
CC   Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de
XX
FH   Key             Location/Qualifiers
FH
FT   source          1. .600
FT                   /db_xref="taxon:9606"
FT                   /organism="Homo sapiens"
FT                   /clone="DKFZp434E1817"
FT                   /clone_lib="434 (synonym: htes3). Vector pSport1; host
FT                   DH10B; sites NotI + SalI"
FT                   /dev_stage="adult"
FT                   /tissue_type="testis"
XX
SQ   Sequence 600 BP; 208 A; 106 C; 119 G; 167 T; 0 other;


alignment_scores:
            Quality:   82.00              Length:    113
              Ratio:  1.123                Gaps:      8
 Percent Similarity:  64.602   Percent Identity: 32.743

alignment_block:
 US-09-068-804-12 x HSM005936   ..

 Align seg 1/1  to: HSM005936  from: 1  to: 600

     118 LeuSerCysGlyGluLysArgTyrThrIleProValLeuHisValLysAs 134
         |||   :::||||||:::|||   ||||||:::   :::|||   :::::
     120 CTAAAAAGTGGTGAAACCAGA...ACAATATCTCACTTTCATTATACTAC 166

     134 nTrpProAspHis...GlnProLeuProSerThrAspGlnLeuGluTyrL 150
         :|||||||||         |||   ::::::::::::   |||:::::::|
     167 CTGGCCAGATTTTGGAGTCCCTGAATCACCAGCTTCATTTCTCAATTTCT 216

     150 euAlaAspArgValLysAsnSerAsnGlnAsnGlyAlaProGlyArgSer 166
         ||      :::|||:::::::|||                   |||
     217 TGTTT...AAAGTGAGAGAATCT.................GGCTCCTTG 245

     167 SerSerAspLysHisLeuProMetIleHisCysLeuGlyGlyValGlyAr 183
         :::::::|||         ::::::||||||||||   :::|||:::|||||
     246 AACCCTGACCATGGGCCTGCGGTGATCCACTGTAGTGCAGGCATTGGGCG 295

     183 gThrGlyThr.........MetAlaAlaAlaLeuValLeuLysAspAsnP 197
         |:::|||||         :::::::::   |||||||||:::::::::
     296 CTCTGGCACCTTCTCTCTGGTAGACACTTGTCTTGTTTTGATGGAAAAAG 345

     197 roHisSer...AsnLeuGluGlnValArgAlaAspPheArgIleHis... 211
```

```
         ::::::   |||:::   ||||||      :::   |||   :::
     346 GAGATGATATTAACATAAAACAAGTGTTACTGAACATGAGAAAATACCGA 395

     212 ...GlyThrIleGluCysTrpLysThrProAlaGlnPhe 223
            |||   |||:::         ||||||:::|||:::
     396 ATGGGTCTTATTCAG.........ACCCCAGATCAACTG 425

seq_name: gb_est30:AI638108

seq_documentation_block:
LOCUS       AI638108      466 bp    mRNA            EST       27-APR-1999
DEFINITION  tt24d10.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2241715 3'
            similar to gb:M68941 PROTEIN-TYROSINE PHOSPHATASE PTP-MEG (HUMAN);,
            mRNA sequence.
ACCESSION   AI638108
NID         g4690342
VERSION     AI638108.1  GI:4690342
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 466)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Jun 22, 1998 this sequence version replaced gi:3247275.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            R. Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco
            High quality sequence stop: 365.
FEATURES             Location/Qualifiers
     source          1. .466
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2241715"
                     /clone_lib="NCI_CGAP_GC6"
                     /tissue_type="pooled germ cell tumors"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Plasmid DNA from the normalized library
                     NCI_CGAP_GC4 was prepared, and ss circles were made in
                     vitro. Following HAP purification, this DNA was used as
                     tracer in a subtractive hybridization reaction. The driver
                     was PCR-amplified cDNAs from a pool of 5,000 clones made
                     from the same library (cloneIDs 1257096-1258631,
                     1469064-1470983, and 1475592-1476743). Subtraction by
                     Bento Soares and M. Fatima Bonaldo. "
BASE COUNT       119 a     98 c     92 g    156 t      1 others
ORIGIN


alignment_scores:
            Quality:   76.50              Length:    117
              Ratio:  1.109                Gaps:      9
 Percent Similarity:  58.974   Percent Identity: 29.915

alignment_block:
 US-09-068-804-12 x AI638108/rev   ..
```

```
Align seg 1/1  to reverse of: AI638108  from: 1  to: 466

    115 AsnMetGlnLeuSerCysGlyGluLysArgTyrThrIleProValLeuHi 131
        :::|||   |||        ||||||      :::   |||:::   ::
    408 AAGATGACCCTATTTAACCAAGAGAAAAATGANAGTCGTCCACTCACTCA 359

    131 sValLys......AsnTrpProAspHis...GlnProLeuProSerThrA 145
        :::::::      :::|||||||||||||      |||      |||:::|
    358 GATCCAGTACATAGCCTGGCCTGACCATGGAGTCCCTGATGATTCGAGTG 309

    145 spGlnLeuGluTyrLeuAlaAspArgValLysAsnSerAsnGlnAsnGly 161
        ||   |||:::::::::      :::|||:::|||
    308 ACTTTCTAGATTTTGTTTGT...CATGTACGAAACAAG..... 274

    162 AlaProGlyArgSerSerSerAspLysHisLeuProMetIleHisCysLe 178
                |||:::::::   :::      :::   ::::::||||||
    273 .........AGGGCTGGCAAGGAAGAACCCGTT...GTTGTCCATTGCAG 236

    178 uGlyGlyValGlyArgThrGly.........ThrMetAlaAlaAlaLeuV 192
         :::|||:::||||||||||||         ||||||::::::|||:::
    235 TGCTGGAATCGGAAGAACTGGGGTTCTTATTACTATGGAAACAGCCATGT 186

    192 alLeuLysAsp......AsnProHisSerAsnLeuGluGlnValArgAla 206
          |||   :::      :::|||         |||:::   ||||||:::
    185 GTCTCATTGAATGCAATCAGCCAGTTTATCCACTAGATATTGTAAGAACA 136

    207 AspPheArgIleHisGlyThrIleGluCysTrpLysThrProAlaGlnPh 223
            |||   :::   ::::::         :::||||||:::|||::
    135 ...ATGAGAGATCAGCGAGCCATGATGATC...CAAACACCTAGTCAATA 92

    223 e 223
        :
     91 C 91

seq_name: gb_est11:AA232769

seq_documentation_block:
LOCUS       AA232769      466 bp    mRNA            EST       06-AUG-1997
DEFINITION  zr45f04.r1 Soares_NhHMPu_S1 Homo sapiens cDNA clone IMAGE:666367 5'
            similar to gb:M68941 PROTEIN-TYROSINE PHOSPHATASE PTP-MEG (HUMAN);,
            mRNA sequence.
ACCESSION   AA232769
NID         g1855780
VERSION     AA232769.1  GI:1855780
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 466)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J.,
            Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-Merck EST Project 1997
  JOURNAL   Unpublished (1997)
COMMENT     On May 5, 1995 this sequence version replaced gi:797826.

            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 570   Std Error: 0.00
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 226.
FEATURES             Location/Qualifiers
     source          1. .466
                     /organism="Homo sapiens"
                     /db_xref="GDB:5428399"
                     /db_xref="taxon:9606"
                     /map="1"
                     /clone="IMAGE:666367"
                     /clone_lib="Soares_NhHMPu_S1"
                     /tissue_type="Pooled human melanocyte, fetal heart, and
                     pregnant uterus"
                     /lab_host="DH10B"
                     /note="Organ: mixed (see below); Vector: pT7T3D-Pac
                     (Pharmacia) with a modified polylinker; Site_1: Not I;
                     Site_2: Eco RI; Equal amounts of plasmid DNA from three
                     normalized libraries (melanocyte 2NbHM, pregnant uterus
                     NbHPU, and fetal heart NbHH19W) were mixed, and ss circles
                     were made in vitro. Following HAP purification, this DNA
                     was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from pools of
                     5,000 clones made from the same 3 libraries. The pools
                     consisted of I.M.A.G.E. clones 260232-265223,
                     340488-345479, and 484488-489479."
BASE COUNT      154 a     86 c    101 g    125 t
ORIGIN


alignment_scores:
           Quality:   75.50            Length:    117
             Ratio:  1.094              Gaps:      9
 Percent Similarity:  58.974   Percent Identity: 29.915

alignment_block:
 US-09-068-804-12 x AA232769   ..

 Align seg 1/1  to: AA232769  from: 1  to: 466

     115 AsnMetGlnLeuSerCysGlyGluLysArgTyrThrIleProValLeuHi 131
         :::|||   |||        ||||||      :::   |||:::   ::
      31 AAGATGACCCTATTTAACCAAGAGAAAAATGAAAGTCGTCCACTCACTCA 80

     131 sValLys......AsnTrpProAspHis...GlnProLeuProSerThrA 145
         :::::::      :::|||||||||||||      |||      |||:::|
      81 GATCCAGTACATAGCCTGGCCTGACCATGGAGTCCCTGATGATTCGAGTG 130

     145 spGlnLeuGluTyrLeuAlaAspArgValLysAsnSerAsnGlnAsnGly 161
         ||   |||:::::::::      :::|||:::|||
     131 ACTTTCTAGATTTTGTTTGT...CATGTACGAAACAAG..... 165

     162 AlaProGlyArgSerSerSerAspLysHisLeuProMetIleHisCysLe 178
                 |||:::::::   :::      :::   ::::::||||||
     166 .........AGGGCTGGCAAGGAAGAACCCGTT...GTTGTCCATTGCAG 203

     178 uGlyGlyValGlyArgThrGly.........ThrMetAlaAlaAlaLeuV 192
          :::|||:::||||||||||||         ||||||::::::|||:::
     204 TGCTGGAATCGGAAGAACTGGGGTTCTTATTACTATGGAAACAGCCATGT 253

     192 alLeuLysAsp......AsnProHisSerAsnLeuGluGlnValArgAla 206
           |||   :::      :::|||         |||:::   ||||||:::
     254 GTCTCATTGAATGCAATCAGCCAGTTTATCCACTAGATATTGTAAGAACA 303

     207 AspPheArgIleHisGlyThrIleGluCysTrpLysThrProAlaGlnPh 223
             |||   :::   ::::::         :::||||||:::|||::
     304 ...ATGAGAGATCAGCGAGCCATGATGATC...CAAACACCTAGTCAATA 347

     223 e 223
         :
     348 C 348

seq_name: gb_est26:AI324939

seq_documentation_block:
LOCUS       AI324939      594 bp    mRNA            EST       23-DEC-1998
DEFINITION  mb50h06.x1 Soares mouse p3NMF19.5 Mus musculus cDNA clone
            IMAGE:332891 3' similar to gb:D37801 Mouse mRNA for protein
            tyrosine phosphatase (MOUSE);, mRNA sequence.
ACCESSION   AI324939
NID         g4059368
```

```
VERSION     AI324939.1  GI:4059368
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 594)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Jan 14, 1998 this sequence version replaced gi:1797466.

            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:214291
            This clone was previously sequenced on the 5' end only, this new
            data is from the 3' end
            High quality sequence stop: 416.
FEATURES             Location/Qualifiers
     source          1. .594
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /map="1; 13q12"
                     /clone="IMAGE:332891"
                     /clone_lib="Soares mouse p3NMF19.5"
                     /dev_stage="19.5 dpc total fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Vector: pT7T3D (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by Bento
                     Soares and M.Fatima Bonaldo. RNA was kindly provided by
                     Dr. Minoru Ko (Wayne State University)."
BASE COUNT      136 a    158 c    144 g    156 t
ORIGIN


alignment_scores:
           Quality:   76.00            Length:    106
             Ratio:  1.267              Gaps:      5
 Percent Similarity:  56.604   Percent Identity: 26.415

alignment_block:
 US-09-068-804-12 x AI324939/rev   ..

 Align seg 1/1  to reverse of: AI324939  from: 1  to: 594

     117 GlnLeuSerCysGlyGluLysArgTyrThrIleProValLeuHisValLy 133
         :::|||      |||:::   |||   |||:::      |||:::   ::
     436 CACCTTCTCACAGGGCAGGAGAGA...ACCGTTTGGCATCTTCAGTACAC 390

     133 sAsnTrpProAspHisGln...ProLeuProSerThrAspGlnLeuGluT 149
         ::::||||||:::|||      |||         ::::::   |||   |
     389 AGACTGGCCTGAACACGGCTGTCCCGAAGACCTCAAAGGATTTTTATCCT 340

     149 yrLeuAlaAspArgValLysAsnSerAsnGlnAsnGlyAlaProGlyArg 165
         |||||::::::                  ::::::::::::::::::

     339 ACCTTGAAGAGATCCAGTCAGTTCGACGTCATACAAACAGTACCAGTGAA 290

     166 SerSerSerAspLysHisLeuProMetIleHisCysLeuGlyGlyValGl 182
         :::   |||::::::         :::::|||||   :::||||||||
     289 CCCAAAAGCCACAATCCCCCGCTGCTGGTGCACTGCAGTGCTGGCGTGGG 240

     182 yArgThrGlyThrMetAlaAlaAlaLeuValLeuLysAspAsnProHisS 199
         ||||||||||           ::::::::::|||  :::         :
     239 AAGAACTGGG.....GTCGTCATTCTGTCAGAGATCATGGTTG 202

     199 erAsnLeuGluGln.ValArgAlaAspPheArgIleHisGlyThrIleGl 215
         ::   ||||||::: :::|||
     201 CCTGTCTGGAACACAATGAGG..... 181

     215 uCysTrpLysThrPro 220
          ||||||::::::|||
     180 .TGCTGGACATCCCCA 166

seq_name: gb_est27:AI431844

seq_documentation_block:
LOCUS       AI431844      539 bp    mRNA            EST       30-MAR-1999
DEFINITION  ti25h05.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2131545 3'
            similar to gb:M68941 PROTEIN-TYROSINE PHOSPHATASE PTP-MEG (HUMAN);,
            mRNA sequence.
ACCESSION   AI431844
NID         g4306026
VERSION     AI431844.1  GI:4306026
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 539)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Mar 10, 1998 this sequence version replaced gi:2948426.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 634   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 392.
FEATURES             Location/Qualifiers
     source          1. .539
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2131545"
                     /clone_lib="NCI_CGAP_Kid11"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Plasmid DNA from the normalized library NCI_CGAP_Kid3 was
                     prepared, and ss circles were made in vitro. Following HAP
                     purification, this DNA was used as tracer in a subtractive
                     hybridization reaction. The driver was PCR-amplified cDNAs
                     from a pool of 5,000 clones made from the same library
                     (cloneIDs 1322376-1323911, 1456007-1456775, and
                     1500552-1502855). Subtraction by Bento Soares and M.
                     Fatima Bonaldo.  "
BASE COUNT      138 a    114 c    108 g    179 t
```

```
ORIGIN


alignment_scores:
            Quality:   75.50              Length:    117
              Ratio:  1.094                Gaps:      9
 Percent Similarity:  58.974   Percent Identity: 29.915

alignment_block:
 US-09-068-804-12 x AI431844/rev   ..

 Align seg 1/1  to reverse of: AI431844  from: 1  to: 539

     115 AsnMetGlnLeuSerCysGlyGluLysArgTyrThrIleProValLeuHi 131
         :::|||   |||         ||||||      :::   |||:::   ::
     408 AAGATGACCCTATTTAACCAAGAGAAAAATGAAAGTCGTCCACTCACTCA 359

     131 sValLys......AsnTrpProAspHis...GlnProLeuProSerThrA 145
         :::::::      :::||||||||||||      |||       |||:::|
     358 GATCCAGTACATAGCCTGGCCTGACCATGGAGTCCCTGATGATTCGAGTG 309

     145 spGlnLeuGluTyrLeuAlaAspArgValLysAsnSerAsnGlnAsnGly 161
         ||   |||:::::::::      :::|||:::|||
     308 ACTTTCTAGATTTTGTTTGT...CATGTACGAAACAAG............ 274

     162 AlaProGlyArgSerSerSerAspLysHisLeuProMetIleHisCysLe 178
                 |||::::::   :::      :::   ::::::||||||
     273 .........AGGGCTGGCAAGGAAGAACCCGTT...GTTGTCCATTGCAG 236

     178 uGlyGlyValGlyArgThrGly.........ThrMetAlaAlaAlaLeuV 192
          :::|||:::|||||||||||||         |||||||::::::|||:::
     235 TGCTGGAATCGGAAGAACTGGGGTTCTTATTACTATGGAAACAGCCATGT 186

     192 alLeuLysAsp......AsnProHisSerAsnLeuGluGlnValArgAla 206
           |||   :::      :::|||         |||:::   |||||||:::
     185 GTCTCATTGAATGCAATCAGCCAGTTTATCCACTAGATATTGTAAGAACA 136

     207 AspPheArgIleHisGlyThrIleGluCysTrpLysThrProAlaGlnPh 223
             |||   :::   ::::::         :::|||||||:::|||::
     135 ...ATGAGAGATCAGCGAGCCATGATGATC...CAAACACCTAGTCAATA 92

     223 e 223
         :
      91 C 91

seq_name: gb_est27:AI442505

seq_documentation_block:
LOCUS       AI442505      375 bp    mRNA            EST       09-MAR-1999
DEFINITION  sa27h06.y1 Gm-c1004 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1004-564 5' similar to TR:Q40938 Q40938 SUCROSE TRANSPORTER. ;,
            mRNA sequence.
ACCESSION   AI442505
NID         g4296649
VERSION     AI442505.1  GI:4296649
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;
            Glycine.
REFERENCE   1  (bases 1 to 375)
  AUTHORS   Shoemaker,R., Briner,D., Marra,M., Hillier,L., Kucaba,T.,
            Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M.,
            Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T.,
            Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S.,
            Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and
            Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     On Mar 10, 1998 this sequence version replaced gi:2948681.
            Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: Genome Systems, Inc. 4633 World
            Parkway Circle St. Louis, Missouri 63134 For further information
            call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
            427-3324 or contact: clones@genomesystems.com or
            info@genomesystems.com web site: www.genomesystems.com
            Seq primer: -40RP from Gibco
            High quality sequence stop: 362.
FEATURES             Location/Qualifiers
     source          1. .375
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1004-564"
                     /clone_lib="Gm-c1004"
                     /tissue_type="root"
                     /lab_host="XL10-Gold"
                     /note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2:
                     XhoI; Root cDNA. The mRNA was isolated from entire roots
                     of 8 day old 'Williams' seedlings which were propagated on
                     paper towels with distilled water. Stratagene's  cDNA
                     Synthesis Kit (catalog #200401) was used to synthesize the
                     cDNA. First- strand synthesis was performed with 5-methyl
                     dCTP, hence the ligated cDNA is hemimethylated.
                     Stratagene's first-strand synthesis primer was used
                     [GAGAGAGAGAGAGAGAGAACTAGTCTCGAG(T)-18]. After
                     second-strand synthesis, the cDNA ends were 'polished'
                     with clone Pfu DNA polymerase, ligated to EcoRI adapters,
                     and phosphorylated. The XhoI site within the first-strand
                     synthesis primer was restricted by digestion with XhoI;
                     all XhoI sites in the  cDNA would be protected by their
                     hemimethylated status. The cDNA constructs were
                     size-fractionated with a 500bp cutoff, using GibcoBRL Life
                     Technologies' cDNA Size Fractionation column. The column
                     eluent was then ligated into  Stratagene's pBluescript II
                     XR Predigested vector (pBluescript II SK(+) that had been
                     digested with EcoRI and XhoI, and phosphorylated). Both
                     the white  and blue colonies appear to contain recombinant
                     plasmids with cDNA inserts. Blue colonies 9n=15) have been
                     sequenced, and possess putative cDNA inserts. This library
                     was constructed by Dr. Paul Keim & Virginia H. Coryell,
                     Department of Biology, Box5640, Northern Arizona
                     University, Flagstaff, AZ 86011, Phone: 520-523-1078 (Dr.
                     Paul Keim), 520-523-1372 (Virginia H. Coryell), Fax:
                     520-523-7500, email: paul.keim@nau.edu,
                     virginia.coryell@nau.edu"
BASE COUNT        57 a     87 c    122 g    109 t
ORIGIN


alignment_scores:
            Quality:   73.50              Length:     90
              Ratio:  1.225                Gaps:       5
 Percent Similarity:  66.667   Percent Identity: 31.111

alignment_block:
 US-09-068-804-12 x AI442505/rev   ..

 Align seg 1/1  to reverse of: AI442505  from: 1  to: 375

     120 CysGlyGluLysArgTyrThrIleProValLeuHisValLysAsnTrpPr 136
         ||||||           ||||||:::   |||:::::            ||
     256 TGTGGTACC.........ACAATTGCAAGATTGAGAACT.........CC 225

     136 oAspHisGlnProLeuProSerThrAspGlnLeuGluTyrLeuAlaAspA 153
         |   :::::|||   |||:::::::::   |||:::|||:::::::::|
     224 CAAAGATAAACCTTGTCCTGCTCCTGAAGCGCTGCAGTATATTGATGCTA 175
```

```
     153 rgValLys..AsnSerAsnGlnAsn...GlyAlaProGlyArgSerSerS 168
         ||:::|||  :::::::::   |||      :::   ::::::::::::
     174 GAGCAAAAGGGACACTGAAAGTAATCGCAAGGGGAACTCCAAGAACCCCA 125

     168 erAspLysHisLeuProMetIleHisCysLeuGlyGlyValGlyArgThr 184
           :::::::|||   |||      |||   |||||||||||||::::::
     124 AAGAAAACCATGGACCCAACCACCACCCCCTCGGAGGGGTGGCCCACGGC 75

     185 GlyThrMetAlaAlaAlaLeuValLeuLys......AspAsnProHisSe 199
         |||   :::   :::||||||||||||||:::      ::::::   |||
      74 GGCGGGGTTCATCCGGCGCTGGTGCTCCGCCACCTTGGTGATCACCACCG 25

     199 rAsnLeuGluGlnValArg 205
          :::   |||:::   |||
      24 TCATTCCGAACCCAATCGC 6

seq_name: gb_est12:AA314946

seq_documentation_block:
LOCUS       AA314946      499 bp    mRNA            EST       19-APR-1997
DEFINITION  EST186775 HCC cell line (matastasis to liver in mouse) II Homo
            sapiens cDNA 5' end similar to similar to tyrosine phosphatase
            CL100, mRNA sequence.
ACCESSION   AA314946
NID         g1967295
VERSION     AA314946.1  GI:1967295
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 499)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
            Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
            White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
            Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
            Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
            Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
            Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
            He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
            Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H.,
            Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
            Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
            Fraser,C.M. and Venter,J.C.
  TITLE     Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   96026280
COMMENT     On Nov 29, 1993 this sequence version replaced gi:430148.
            Other_ESTs: THC171716
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES              Location/Qualifiers
     source           1. .499
                      /organism="Homo sapiens"
                      /db_xref="ATCC (inhost):113107"
                      /db_xref="taxon:9606"
                      /clone_lib="HCC cell line (matastasis to liver in mouse)
                      II"
                      /tissue_type="colon"
                      /cell_type="KM12SM"
                      /cell_line="KM12C(HCC)metastasis into mouse (liver)"
                      /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                      XhoI"
BASE COUNT      114 a    127 c    162 g     95 t      1 others
ORIGIN


alignment_scores:
            Quality:   74.00            Length:     96
              Ratio:  1.345              Gaps:      4
 Percent Similarity:  57.292   Percent Identity: 26.042

alignment_block:
 US-09-068-804-12 x AA314946   ..

 Align seg 1/1  to: AA314946  from: 1  to: 499

     142 ProSerThrAspGlnLeuGluTyrLeuAlaAspArgValLysAsnSerAs 158
         |||   ||||||   |||   :::|||:::      |||       ::
     236 CCCGAGACGGACCTACTCAGCCATCTGGACCGGTGCGTGGCCTTCATCGG 285

     158 nGlnAsnGlyAlaProGlyArgSerSerSerAspLysHisLeuProMetI 175
         :|||:::   |||   |||||||:::                    ::::
     286 TCAGGCCCGCGCTGAGGGCCGTGCGGTG.................TTGG 317

     175 leHisCysLeuGlyGlyValGlyArgThrGlyThrMetAlaAlaAlaLeu 191
         ::||||||   :::|||||||:::|||:::   ::::::   :::|||:::
     318 TGCACTGTCATGCAGGAGTCAGTCGAAGTGTGGCCATAATAACTGCTTTT 367

     192 ValLeuLys.................AspAsnProHisSerAsnLeuGl 202
         ::::::|||                 ::::::::::::   :::|||::
     368 CTCATGAAGACTGACCAACTTCCCTTTGAAAAAGCCTATGAAAAGCTCCA 417

     202 uGlnValArgAlaAspPheArgIleHisGlyThrIleGluCysTrpLysT 219
         :   ::::::::::::   :::::::::      :::|||   |||
     418 GATTCTCAAACCAGAGGCTAAGATGAATGAGGGGTTTGAG...TGG.... 460

     219 hrProAlaGlnPheValGlnLeuLysAlaMetGlnAla 231
                       |||||||||      ||||||
     461 .................CAACTGAAATTATACCAGGCA 481

seq_name: gb_est12:AA296865

seq_documentation_block:
LOCUS       AA296865      356 bp    mRNA            EST       18-APR-1997
DEFINITION  EST112446 Aorta endothelial cells, TNF alpha-treated Homo sapiens
            cDNA 5' end similar to tyrosine phosphatase D1, mRNA sequence.
ACCESSION   AA296865
NID         g1949219
VERSION     AA296865.1  GI:1949219
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 356)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
            Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
            White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
            Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
            Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
            Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
            Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
            He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
            Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H.,
            Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
```

```
            Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
            Fraser,C.M. and Venter,J.C.
  TITLE     Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   96026280
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1290545.
            Other_ESTs: THC167908
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES              Location/Qualifiers
     source           1. .356
                      /organism="Homo sapiens"
                      /db_xref="ATCC (inhost):116524"
                      /db_xref="taxon:9606"
                      /clone_lib="Aorta endothelial cells, TNF alpha-treated"
                      /cell_type="endothelial cell"
                      /dev_stage="adult"
                      /note="Organ: aorta; Vector: pBluescript SK-; Site_1:
                      EcoRI; Site_2: XhoI"
BASE COUNT       89 a     96 c     96 g     73 t      2 others
ORIGIN


alignment_scores:
            Quality:   72.50             Length:    105
              Ratio:  1.272               Gaps:      4
 Percent Similarity:  54.286   Percent Identity: 23.810

alignment_block:
 US-09-068-804-12 x AA296865   ..

 Align seg 1/1  to: AA296865  from: 1  to: 356

     117 GlnLeuSerCysGlyGluLysArgTyrThrIleProValLeuHisValLy 133
         :::|||      |||:::   |||   |||:::      ||:::   ::
      55 CACCTCCTTACTGGGCAAGAGAGG...ACCGTNTGGCACCTCCAATACAC 101

     133 sAsnTrpProAspHisGlnProLeuProSerThrAspGlnLeu..... 147
         ::::||||||:::|||           ::::::::::|||
     102 AGACTGGCCTGAACAT.........GGCTGTCCAGAAGACCTCAAGGGAT 142

     148 .....GluTyrLeuAlaAspArgValLysAsnSerAsnGlnAsnGlyAla 162
               ||||||::::::          ::::::::::
     143 TTTNATCATATCTTGAAGAGATCCAGTCTGTTCGACGCCATACAAATAGC 192

     163 ProGlyArgSerSerSerAspLysHisLeuProMetIleHisCysLeuGl 179
         ::::::   :::   |||   :::         ::::::||||||   ::
     193 ACAAGTGATCCCCAAAGCCCCAACCCTCCGTTGTTGGTCCACTGCAGTGC 242

     179 yGlyValGlyArgThrGlyThrMetAlaAlaAlaLeuValLeuLysAspA 196
         :|||||||||||||||||::::::      :::   :::
     243 TGGGGTAGGAAGGACTGGCGTGGTGATTTTTTCGGAGATCCATGATCGCC 292

     196 snProHisSerAsnLeuGluGlnValArgAlaAspPheArgIleHisGly 212
           |||   :::::::::                  |||
     293 TGCCTGGGAACACAATG.....................AGGTGCTGGACA 321

     213 ThrIleGluCysTrp 217
         :::   |||||||||
     322 TCCCGAGAGTGCTGG 336

seq_name: gb_est17:C70185
```

```
seq_documentation_block:
LOCUS       C70185        360 bp    mRNA            EST       29-SEP-1997
DEFINITION  C70185 Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
            clone yk384b2 5', mRNA sequence.
ACCESSION   C70185
NID         g2440710
VERSION     C70185.1  GI:2440710
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 360)
  AUTHORS   Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,
            Sano,M., Miyata,A. and Nishigaki,A.
  TITLE     Expression map of the C.elegans genome
  JOURNAL   Unpublished (1996)
COMMENT     On May 8, 1995 this sequence version replaced gi:800911.

            Contact: Yuji Kohara
            Gene Library Lab
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411, Japan
            Tel: 0559-75-0771
            Fax: 0559-75-6240
            Email: ykohara@ddbj.nig.ac.j.
FEATURES              Location/Qualifiers
     source           1. .360
                      /organism="Caenorhabditis elegans"
                      /strain="CB1489 him-8(e1489)"
                      /db_xref="taxon:6239"
                      /clone="yk384b2"
                      /clone_lib="Yuji Kohara unpublished cDNA"
                      /note="dev_stage=varied, sex=Hermaphrodite male,
                      tissue_type=whole animal"
BASE COUNT      106 a     73 c     84 g     97 t
ORIGIN


alignment_scores:
            Quality:   72.50             Length:     77
              Ratio:  1.648               Gaps:       2
 Percent Similarity:  57.143   Percent Identity: 28.571

alignment_block:
 US-09-068-804-12 x C70185   ..

 Align seg 1/1  to: C70185  from: 1  to: 360

     135 TrpProAspHisGlnProLeuProSerThrAspGlnLeuGluTyrLeuAl 151
         ||||||||||||!             ::::::   ::::::      :::::
      12 TGGCCCGACCATGGAGTTCCAGACGATCCGAAACACTTTATTCAATTTGT 61

     151 aAspArgValLysAsnSerAsnGlnAsnGlyAlaProGlyArgSerSerS 168
         :|||   |||::::::::   |||:::::::::
      62 GGATGAAGTTCGAAAAGCACGACAAGGATCAGTTGATCCGATT....... 104

     168 erAspLysHisLeuProMetIleHisCysLeuGlyGlyValGlyArgThr 184
                       ::::::||||||   :::|||:::|||||||||
     105 .................GTTGTACACTGTTCGGCGGGAATTGGAAGAACT 137

     185 GlyThrMet..............AlaAlaAlaLeuValLeuLysAspAs 196
         |||::::::               ||||||   ||||||      :::::
     138 GGGGTTCTTATTTTAATGGAGACTGCCGCGTGTCTTGTTGAATCAAATGA 187

     196 nProHisSerAsnLeuGluGlnValArgAla 206
         :|||         |||:::   ||||||:::
     188 ACCTGTATATCCATTAGATATTGTGAGGACG 218

seq_name: gb_est17:AA627031

seq_documentation_block:
LOCUS       AA627031      370 bp    mRNA            EST       16-OCT-1997
```

```
DEFINITION  MBAFCW2G01T3 Brugia malayi adult female cDNA (SAW96MLW-BmAF) Brugia
            malayi cDNA clone AFCW2G01 5', mRNA sequence.
ACCESSION   AA627031
NID         g2539418
VERSION     AA627031.1  GI:2539418
KEYWORDS    EST.
SOURCE      Brugia malayi.
  ORGANISM  Brugia malayi
            Eukaryota; Metazoa; Nematoda; Secernentea; Spiruria; Spirurida;
            Filarioidea; Onchocercidae; Brugia.
REFERENCE   1  (bases 1 to 370)
  AUTHORS   Blaxter,M.L., Waterfall,M., Daub,J., Lizotte,M., Baron,L. and
            Jones,S.J.
  TITLE     Genes expressed in adult female Brugia malayi
  JOURNAL   Unpublished (1996)
COMMENT     On Apr 14, 1993 this sequence version replaced gi:693144.

            Contact: Blaxter ML
            Institute of Cell, Animal and Population Biology
            University of Edinburgh
            Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9
            3JT, UK.
            Tel: +44 131 650 6760
            Fax: +44 131 670 5450
            Email: mark.blaxter@ed.ac.uk
            The ABI trace of this sequence can be viewed at
            http://www.sanger.ac.uk/brugia/AFC/MBAFCW2G01T3.html
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .370
                     /organism="Brugia malayi"
                     /db_xref="taxon:6279"
                     /clone="AFCW2G01"
                     /clone_lib="Brugia malayi adult female cDNA
                     (SAW96MLW-BmAF)"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="XL1-Blue MRF'"
                     /note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:
                     Xho I; Lymphatic filarial nematode parasite of humans.
                     mRNA was prepared from approximately 50 adult females
                     isolated from the  peritoneal cavity of jirds and
                     converted to double-stranded cDNA        using reverse
                     transcriptase and oligo(dT) followed by RNase H and DNA
                     pol I. The library has 5 x 10E6 independent recombinants
                     and the average insert size is ~900bp.The library was
                     constructed by          Michelle Lizotte-Waniewski. The
                     library is available from      Dr.S.A.Williams, email:
                     genome@smith.edu."
BASE COUNT      111 a     68 c     95 g     96 t
ORIGIN


alignment_scores:
           Quality:   72.50            Length:     90
             Ratio:  1.368              Gaps:      4
 Percent Similarity:  58.889   Percent Identity: 27.778

alignment_block:
 US-09-068-804-12 x AA627031   ..

 Align seg 1/1  to: AA627031  from: 1  to: 370

     122 GluLysArgTyrThrIleProValLeuHisValLysAsnTrpProAspHi 138
         |||:::|||      ::::::   ::::::   ::::::||||||||||
     108 GAGAGACGG......GIAAGCCAAATGCAGTACACGGCGTGGCCTGATCA 151

     138 sGlnProLeuProSerThrAspGlnLeuGluTyrLeuAlaAspArgValL 155
         |           ::::::  ::::::      :::::|||   |||:
     152 TGGTGTTCCAGATGATCCGAAACATTTTATTGATTTCGTCGATGAAGTTC 201

     155 ysAsnSerAsnGlnAsnGlyAlaProGlyArgSerSerSerAspLysHis 171
         ::                       :::|||::::::|||

     202 GA.....................CGTGCACGAGCAGGTTCAGTAGATCCA 230

     172 LeuProMetIleHisCysLeuGlyGlyValGlyArgThrGlyThrMet.. 187
         :::   ::::::||||||   :::|||:::|||||||||||||::::::
     231 ATC...GTTGTGCACTGCAGTGCAGGAATCGGACGAACCGGTGTGCTAAT 277

     188 .............AlaAlaAlaLeuValLeuLysAspAsnProHisSerA 200
                       ||||||   ||||||       ::::::|||
     278 ACTGATGGAAACAGCTGCTTGTCTGGTAGAAGGTAATGAACCGGTGTATC 327

     200 snLeuGluGlnValArgAla 206
           |||:::   ||||||:::
     328 CCCTAGATATTGTTAGAACA 347

seq_name: gb_est19:AA805956

seq_documentation_block:
LOCUS       AA805956      334 bp    mRNA            EST       12-FEB-1998
DEFINITION  oc20b03.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1341389 3'
            similar to TR:O00633 O00633 PROTEIN TYROSINE PHOSPHATASE. ;, mRNA
            sequence.
ACCESSION   AA805956
NID         g2874706
VERSION     AA805956.1  GI:2874706
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 334)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 17, 1998 this sequence version replaced gi:1900233.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
            Ph.D., Gerald Marti, M.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 319.
FEATURES             Location/Qualifiers
     source          1. .334
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="21"
                     /clone="IMAGE:1341389"
                     /clone_lib="NCI_CGAP_GCB1"
                     /tissue_type="germinal center B cell"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from human tonsillar cells enriched for
                     germinal center B cells by flow sorting (CD20+, IgD-),
                     provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                     (NCI) and Dr. Gerald Marti (CBER).  cDNA synthesis was
                     primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTT-
                     3']. Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo."
```

```
BASE COUNT      119 a     57 c     68 g     90 t
ORIGIN


alignment_scores:
            Quality:   72.00             Length:     93
              Ratio:  1.469               Gaps:      4
 Percent Similarity:  52.688   Percent Identity: 26.882

alignment_block:
 US-09-068-804-12 x AA805956   ..

 Align seg 1/1  to: AA805956  from: 1  to: 334

     112 AspGlnTyrAsnMetGlnLeuSerCysGlyGluLysArgTyrThrIlePr 128
         :::::::|||:::::::         |||:::|||:::::::|||   :::::
      10 AACCATTACAAGATATACAATCTTTGTGCTGAAAGACATTATGACACCGC 59

     128 oValLeuHisValLys.............AsnTrpProAspHisGlnProL 141
         :   ::::::   :::                          |||||||:::|||
      60 CAAATTTAATTGCAGAGTTGCACAATATCCTTTTGAAGACCATAACCCA. 108

     141 euProSerThrAspGlnLeuGluTyr.....................Leu 150
                      |||||||||                          |||
     109 ...........CCACAGCTAGAACTTATCAAACCCTTTTGTGAAGATCTT 147

     151 AlaAspArgValLysAsnSerAsnGlnAsnGlyAlaProGlyArgSerSe 167
         ::::::::::::                                   |||
     148 GACCAATGGCTA.................................AGTGA 164

     167 rSerAspLysHisLeuProMetIleHisCysLeuGlyGlyValGlyArgT 184
          :::|||:::|||::::::   |||||||||   :::|||   |||||||
     165 AGATGACAATCATGTTGCAGCAATTCACTGTAAAGCTGGAAAGGGACGAA 214

     184 hrGlyThrMetAlaAlaAlaLeuValLeu 193
         |||||:::|||      |||   :::|||
     215 CTGGTGTAATGATATGTGCATATTTATTA 243

seq_name: gb_est11:AA257490

seq_documentation_block:
LOCUS       AA257490      480 bp    mRNA            EST       17-MAR-1997
DEFINITION  MBAFCZ0E10T3 Brugia malayi adult female cDNA (SAW96MLW-BmAF) Brugia
            malayi cDNA clone AFCZ0E10 5', mRNA sequence.
ACCESSION   AA257490
NID         g1892966
VERSION     AA257490.1  GI:1892966
KEYWORDS    EST.
SOURCE      Brugia malayi.
  ORGANISM  Brugia malayi
            Eukaryota; Metazoa; Nematoda; Secernentea; Spiruria; Spirurida;
            Filarioidea; Onchocercidae; Brugia.
REFERENCE   1  (bases 1 to 480)
  AUTHORS   Blaxter,M.L., Waterfall,M., Daub,J., Lizotte,M., Baron,L. and
            Jones,S.J.
  TITLE     Genes expressed in adult female Brugia malayi
  JOURNAL   Unpublished (1996)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1397518.

            Contact: Blaxter ML
            Institute of Cell, Animal and Population Biology
            University of Edinburgh
            Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9
            3JT, UK.
            Tel: +44 131 650 6760
            Fax: +44 131 670 5450
            Email: mark.blaxter@ed.ac.uk
            The ABI trace of this sequence can be viewed at
            http://www.sanger.ac.uk/brugia/AFC/MBAFCZ0E10T3.html
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .480
                     /organism="Brugia malayi"
                     /db_xref="taxon:6279"
                     /map="X q21-q24; 17q21"
                     /clone="AFCZ0E10"
                     /clone_lib="Brugia malayi adult female cDNA
                     (SAW96MLW-BmAF)"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="XL1-Blue MRF'"
                     /note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:
                     Xho I; Lymphatic filarial nematode parasite of humans.
                     mRNA was prepared from approximately 50 adult females
                     isolated from the  peritoneal cavity of jirds and
                     converted to double-stranded cDNA        using reverse
                     transcriptase and oligo(dT) followed by RNase H and DNA
                     pol I. The library has 5 x 10E6 independent recombinants
                     and the average insert size is ~900bp.The library was
                     constructed by          Michelle Lizotte-Waniewski. The
                     library is available from      Dr.S.A.Williams, email:
                     genome@smith.edu."
BASE COUNT      131 a     88 c    111 g    150 t
ORIGIN


alignment_scores:
            Quality:   73.50             Length:     89
              Ratio:  1.413               Gaps:      3
 Percent Similarity:  58.427   Percent Identity: 28.090

alignment_block:
 US-09-068-804-12 x AA257490   ..

 Align seg 1/1  to: AA257490  from: 1  to: 480

     123 LysArgTyrThrIleProValLeuHisValLysAsnTrpProAspHisGl 139
         :::|||   |||   :::   ::::::   ::::::|||||||||||||
      27 AGAAGAGAGACGGGTAGCCAAATGCAGTACACGGCGTGGCCTGATCATGG 76

     139 nProLeuProSerThrAspGlnLeuGluTyrLeuAlaAspArgValLysA 156
                 ::::::   ::::::      ::::::|||   |||:::
      77 TGTTCCAGATGATCCGAAACATTTTATTGATTTCGTCGATGAAGTTCGA. 125

     156 snSerAsnGlnAsnGlyAlaProGlyArgSerSerSerAspLysHisLeu 172
                              :::|||::::::|||         :::
     126 ....................CGTGCACGAGCAGGTTCAGTAGATCCAATC 155

     173 ProMetIleHisCysLeuGlyGlyValGlyArgThrGlyThrMet..... 187
         ::::::||||||   :::|||:::||||||||||||||::::::
     156 ...GTTGTGCACTGCAGTGCAGGAATCGGACGAACCGGTGTGCTAATACT 202

     188 .........AlaAlaAlaLeuValLeuLysAspAsnProHisSerAsnL 201
                 ||||||   ||||||      ::::::|||         |
     203 GATGGAAACAGCTGCTTGTCTGGTAGAAGGTAATGAACCGGTGTATCCCC 252

     201 euGluGlnValArgAla 206
         ||:::   ||||||:::
     253 TAGATATTGTTAGAACA 269

seq_name: gb_est22:AA995560

seq_documentation_block:
LOCUS       AA995560      488 bp    mRNA            EST       27-JUL-1998
DEFINITION  os02a12.s1 NCI_CGAP_Lu5 Homo sapiens cDNA clone IMAGE:1604158 3'
            similar to gb:D13540 PROTEIN-TYROSINE PHOSPHATASE 2C (HUMAN);, mRNA
            sequence.
ACCESSION   AA995560
NID         g3182049
VERSION     AA995560.1  GI:3182049
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 488)
```

```
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 14, 1998 this sequence version replaced gi:1796911.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 716   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 318.
FEATURES              Location/Qualifiers
     source           1. .488
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /map="9-12"
                      /clone="IMAGE:1604158"
                      /clone_lib="NCI_CGAP_Lu5"
                      /tissue_type="carcinoid"
                      /lab_host="DH10B"
                      /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                      modified polylinker; 1st strand cDNA was prepared from
                      neuroendocrine lung carcinoid, and was then primed with a
                      Not I - oligo(dT) primer. Double-stranded cDNA was ligated
                      to Eco RI adaptors (Pharmacia), digested with Not I and
                      cloned into the Not I and Eco RI sites of the modified
                      pT7T3 vector.  Library is normalized. Library was
                      constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      107 a    145 c    101 g    135 t
ORIGIN


alignment_scores:
            Quality:   73.50              Length:    155
              Ratio:  0.993                Gaps:      8
 Percent Similarity:  47.742   Percent Identity: 23.226

alignment_block:
 US-09-068-804-12 x AA995560/rev   ..

 Align seg 1/1  to reverse of: AA995560  from: 1  to: 488

      49 ArgProAspArgLeuGlu.....AlaHisMetLysMetLeuLe 61
         |||   :::|||::::::           :::   |||   ||||||
     435 AGAGAGGAAAGAGTAAATGTGTCAAATACTGGCCTGATGAGTATGCTCTA 386

      61 uGluLysGluCysSerCysLeuValValLeuThrSerGluArgSerAspA 78
            :::      ||||||::::::      :::     |||
     385 AAAGAATATGGCGTCATGCGTGTTAGGAACGTCAAGAAAGCGC..... 343

      78 laGlyLysThrIleThrThrTyrPheArgGly.SerTyrThrPheGlyGl 94
                                 |||::: :::|||||::   ||
     342 .........................CGCTCATGACTATACGCTAAGAGA 319

      94 uValHisThrAsnSerGlnLysValSerSerAlaSerGlnGlyGluAlaI 111
         |:::      :::      ||||||:::   ::::::        ::::
     318 ACTTAAACTTTCA......AAGGTTGGACAAGGGAATACGGAGAGAACGG 275

     111 leAspGlnTyrAsnMetGlnLeuSerCysGlyGluLysArgTyrThrIle 127
         ::   ||||||
     274 TCTGGCAATAC..................................... 264

     128 ProValLeuHisValLysAsnTrpProAspHis...GlnProLeuProSe 143

                 |||   ::::::||||||||||||||       |||      ::
     263 .........CACTTTCGGACCTGGCCGGACCACGGCGTGCCCAGCGACCC 223

     143 rThrAspGlnLeuGluTyrLeuAlaAsp......ArgValLysAsnSerA 158
         :   :::   |||::::::|||::::::      :::   :::::::|||
     222 TGGGGGCGTGCTGGACTTCCTGGAGGAGGTGCACCATAAGCAGGAGAGCA 173

     158 snGlnAsnGlyAlaProGlyArgSerSerSerAspLysHisLeuProMet 174
               :::::::::|||                                :::
     172 TCATGGATGCAGGGCCG.........................GTCGTG 150

     175 IleHisCysLeuGlyGlyValGlyArgThrGlyThrMetAlaAlaAlaLe 191
         :::||||||   :::|||:::||||||||||||||||      :::
     149 GTGCACTGCAGTGCTGGAATTGGCCGGACAGGGACGTTCATTGTGATTGA 100

     191 uValLeuLysAsp 195
          :::|||   |||
      99 TATTCTTATTGAC 87
```

OM of: US-09-068-804-12 to: Issued_Patents_NA:* out_format : pfs

Date: Oct 5, 1999 7:12 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=x1p
-Q=/cgn2_1/USPTO_spool/US09068804/runat_05101999_145135_10326/app_query.fasta.1
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=pam150
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=escore
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09068804
-NCPU=6 -ICPU=3 -WAIT -THREADS=1

Search information block:
Query: US-09-068-804-12
Query length: 239
Database: Issued_Patents_NA:*
Database sequences: 192659
Database length: 52021692
Search time (sec): 101.650000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ..
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-586-272-1	-	110.00	235.06	1.1e-05	2406	!
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-342-930-1	+	92.00	180.52	0.0115	5455	!
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-652-971-1	+	89.00	172.33	0.0329	5769	!
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-991-258A-1	+	89.00	172.33	0.0329	5769	!
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-202-389-11	+	84.50	171.36	0.0373	2276	!
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-018-129-4	+	84.00	167.82	0.0586	2790	!
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-791-522-2	+	79.00	167.54	0.0608	924	! S
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-202-389-7	+	82.50	166.98	0.0653	2143	!
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-201-697-3	+	84.50	163.14	0.1068	4752	!
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-201-697-6	+	84.50	163.03	0.1083	4799	!
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-201-697-5	+	84.50	163.00	0.1089	4815	!
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-015-985-4	+	81.50	161.19	0.1372	2872	!
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-015-985-2	+	80.50	160.63	0.1474	2409	!
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-348-006B-6	+	84.50	160.54	0.1491	6000	!
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-800-825A-6	+	84.50	160.54	0.1491	6000	!
/cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:PCT-US94-10166-6	+	84.50	160.54	0.1491	6000	
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-202-389-5	+	78.00	155.62	0.2803	2145	!
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-446-345-33	+	72.50	151.71	0.4627	878	!
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-201-697-1	+	80.50	151.06	0.5028	5679	!
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-202-389-1	+	73.50	146.56	0.8961	1747	!
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-616-844-7	+	78.50	144.67	1.14	6407	!
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-599-654-7	+	78.50	144.67	1.14	6407	!
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-202-389-9	+	66.00	144.66	1.14	380	! S
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-336-618-25	+	73.50	144.21	1.21	2157	!
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-036-210-8	+	66.50	142.17	1.57	532	! S
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-449-609-8	+	66.50	142.17	1.57	532	! S
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-400-115-3	+	68.00	141.14	1.80	819	! S
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-006-676B-2	+	74.00	140.66	1.91	3319	!
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-282-845-1	+	74.00	140.66	1.91	3319	!
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-428-414A-4	+	74.00	140.66	1.91	3319	!
/cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:PCT-US94-00324-2	+	74.00	140.66	1.91	3319	
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-447-464-2	+	75.00	137.17	2.99	5690	!
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-716-679-2	+	75.00	137.17	2.99	5690	!
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-596-291-1	+	76.50	137.10	3.02	8040	!
/cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:PCT-US94-09943-1	+	76.50	137.09	3.02	8043	
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-933-750C-70	+	66.00	136.07	3.44	821	!
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-449-644-3	+	73.00	135.06	3.92	4374	!
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-087-244A-3	+	73.00	135.06	3.92	4374	!
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-449-644-4	+	73.00	134.37	4.28	4651	!
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-087-244A-4	+	73.00	134.37	4.28	4651	!
/cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:PCT-US95-05512-1	+	73.00	133.31	4.90	5117	
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-239-276-9	+	66.50	131.27	6.37	1413	!
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-446-345-35	+	70.50	129.53	7.96	4080	
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-036-210-10	+	66.50	125.79	12.86	2309	
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-449-609-10	+	66.50	125.79	12.86	2309	

seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-586-272-1

seq_documentation_block:
; Sequence 1, Application US/08586272
; Patent No. 5824795
; GENERAL INFORMATION:
; APPLICANT: POPOFF, MICHEL Y.
; APPLICANT: LE GUERN FELLOUS, MURIEL
; TITLE OF INVENTION: OLIGONUCLEOTIDES FOR THE DETECTION OF
; TITLE OF INVENTION: SALMONELLA
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/586,272
; FILING DATE: 16-JAN-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 9500410
; FILING DATE: 16-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 660-107-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2406 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: SALMONELLA SER. TYPHI
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 97..1755
; OTHER INFORMATION: /product= "iagA"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1776..2255
; OTHER INFORMATION: /product= "iagB"
US-08-586-272-1

alignment_scores:
Quality: 110.00 Length: 31
Ratio: 3.793 Gaps: 1
Percent Similarity: 93.548 Percent Identity: 93.548

alignment_block:
US-09-068-804-12 x US-08-586-272-1/rev ..

Align seg 1/1 to reverse of: US-08-586-272-1 from: 1 to: 2406

210 IleHisGlyThrIleGluCysTrpLysThrProAlaGlnPheValGlnLe 226

```
          |||||||||||||||||||||||||||||||||   ||||||||||||||
    2404 ATTCACGGAACAATAGAATGCTGGAAGACGCCTCA.CAGTTTGTACAACT 2356

      226 uLysAlaMetGlnAla.GlnLeuLeuMetThrThrAlaSer 239
          ||||||||||||||| ||||||||||||||||||||||||
    2355 AAAAGCAATGCAAGCGCCAGTTGCTTATGACGACGGCAAGC 2315

seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-342-930-1

seq_documentation_block:
; Sequence 1, Application US/08342930
; Patent No. 5821084
;  GENERAL INFORMATION:
;    APPLICANT:  OLMSTED, ELIZABETH A.
;    APPLICANT:  MAURO, LAURA J.
;    APPLICANT:  DAVIS, ALAN R.
;    APPLICANT:  DIXON, JACK E.
;    TITLE OF INVENTION:  OSTEOBLAST-TESTICULAR PROTEIN TYROSINE
;    TITLE OF INVENTION:  PHOSPHATASE
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  755 Page Mill Road
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304-1018
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/342,930
;      FILING DATE:  21-NOV-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KONSKI, ANTOINETTE F.
;      REGISTRATION NUMBER:  34,202
;      REFERENCE/DOCKET NUMBER:  20344-20975.00
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 813-5600
;      TELEFAX:  (415) 494-0792
;      TELEX:  706141
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5455 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  205..5337
US-08-342-930-1


alignment_scores:
            Quality:   92.00              Length:    214
              Ratio:  0.800                Gaps:      9
 Percent Similarity:  53.738   Percent Identity: 20.093

alignment_block:
 US-09-068-804-12 x US-08-342-930-1   ..

 Align seg 1/1  to: US-08-342-930-1  from: 1  to: 5455

      17 GlnValLysLeuSerAspGlyMetProValProValIleArg..... 30
         |||   :::|||   :::      |||      :::|||:::
    3694 CAGCCCCGACTAGAGGCTGAGCATCCGGACAACATCATCAAGAACCGGTA 3743

      31 .........LeuThrPheAspGlyLysProValAlaLeuAlaGlyIle. 43
                  |||:::::|||        |||   |||:::   :::
    3744 CCCACACGTGCTGCCCTATGACCACTCCAGGGTCAGGCTGACCCAGCTAC 3793

      43 ........................................... 43

    3794 CAGGAGAGCCTCATTCTGACTACATCAATGCCAACTTCATCCCAGGCTAT 3843

      44 ...ArgThrGlnLysIleArgProAspArg.................Le 53
           :::|||||  |||   :::  :::                    ||
    3844 AGCCACACACAGGAGATCATTGCCACCCAGGGGCCTCTCAAAAAGACGCT 3893

      53 uGluAlaHisMetLysMetLeuLeuGluLysGluCysSerCysLeuValV 70
         ||||:::     :::::::::  |||::::::          :::::::
    3894 AGAGGACTTCTGGCGGTTGGTATGGGAGCAGCAAGTCCACGTGATCATCA 3943

      70 alLeuThrSerGluArgSerAspAlaGlyLysThrIleThrThrTyrPhe 86
         ::|||||            :::::::|||:::::::::      ::::::
    3944 TGCTGACTGTGGGCATG...GAGAACGGGCGGGTACTGTGTGAGCACTAC 3990

      87 Arg......GlySerTyrThrPheGlyGluValHisThrAsnSerGlnLy 101
         :::      :::|||  :::      :::   |||::::::   :::
    3991 TGGCCAGCCAACTCCACGCCTGTTACTCACGGTCACATCACCATCCACCT 4040

     101 sValSerSerAlaSerGlnGlyGlu...AlaIleAspGlnTyrAsnMetG 117
          ::::::   :::::::::::|||   :::      ::::::::::::|
    4041 CCTGGCAGAGGAGCCTGAGGATGAGTGGACCAGGAGGGAATTCCAGCTGC 4090

     117 lnLeuSerCysGlyGluLysArgTyrThrIleProValLeuHisValLys 133
         ||   :::      :::|||:::      :::      |||:::   :::
    4091 AGCACGGTACCGAGCAAAAACAGAGGCGAGTGAAGCAGCTGCAGTTCACT 4140

     134 AsnTrpProAspHisGln...ProLeuProSerThrAspGlnLeuGluTy 149
         :::|||||||||||||      |||   :::::::::::   |||:::::
    4141 ACCTGGCCAGACCACAGTGTCCCGGAGGCTCCCAGCTCTCTGCTCGCTTT 4190

     149 rLeuAla.........AspArgValLysAsnSerAsnGlnAsnGlyAlaP 163
         :::::::         :::::::|||:::::::::::   :::|||:::
    4191 TGTAGAACTGGTACAGGAGCAGGTGCAGGCCACTCAGGGCAAGGGACCC. 4239

     163 roGlyArgSerSerSerAspLysHisLeuProMetIleHisCysLeuGly 179
                                          ::::::||||||   :::
    4240 ............................ATCCTGGTGCATTGCAGTGCT 4260

     180 GlyValGlyArgThrGlyThrMetAlaAlaAlaLeuValLeu 193
         ||||||||||||||||||||   :::|||   |||   |||
    4261 GGCGTGGGGAGGACAGGCACCTTTGTGGCTCTCTTGCGGCTA 4302

seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-652-971-1

seq_documentation_block:
; Sequence 1, Application US/08652971
; Patent No. 5814507
;  GENERAL INFORMATION:
;    APPLICANT:  Cheng, Jill
;    APPLICANT:  Lasky, Laurence A.
;    TITLE OF INVENTION:  A NOVEL KAPPA/MU-LIKE PROTEIN TYROSINE
;    TITLE OF INVENTION:  PHOSPHATASE, PTP LAMBDA
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd.
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/652,971
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
```

```
;       NAME:  Dreger, Ginger R.
;       REGISTRATION NUMBER:  33,055
;       REFERENCE/DOCKET NUMBER:  P1033
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415) 225-3216
;       TELEFAX:  (415) 952-9881
;       TELEX:  910 371-7168
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  5769 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  379..4686
US-08-652-971-1


alignment_scores:
            Quality:   89.00              Length:    151
              Ratio:  1.011                Gaps:      7
 Percent Similarity:  58.278   Percent Identity: 23.179

alignment_block:
 US-09-068-804-12 x US-08-652-971-1   ..

 Align seg 1/1  to: US-08-652-971-1  from: 1  to: 5769

      43 IleArgThrGlnLysIleArgProAspArgLeuGluAlaHisMetLysMe 59
         |||   ||||||      :::|||:::   :::   :::      :::|| 
    3223 ATAGCCACTCAAGGGCCAAAGCCTGAGATGATCTACGATTTCTGGCGCAT 3272

      59 tLeuLeuGluLysGluCysSerCysLeuValValLeuThrSerGluArgS 76
         |:::   :::   :::|||:::::   :::|||:::::::::
    3273 GGTGTGGCAGGAACAGTGTGCGAGC...ATCGTCATGATCACCAAGCTGG 3319

      76 erAspAlaGlyLysThrIleThrThrTyrPheArg......GlySerTyr 90
           ::::::|||:::::      :::   :::::      :::|||
    3320 TAGAGGTGGGCAGGGTGAAGTGTTCTCGCTACTGGCCTGAGGACTCAGAC 3369

      91 ThrPheGlyGluValHisThrAsnSerGlnLysValSerSerAlaSerGl 107
           :::|||:::::   :::::      |||:::
    3370 ATGTATGGGGACATCAAGATCACGCTGGTAAAGACA.............. 3405

     107 nGlyGluAlaIleAspGlnTyrAsnMetGlnLeuSerCysGlyGluLys. 123
           |||:::::::::::::|||   :::::           |||:::
    3406 ....GAGACACTGGCTGAGTATGTGGTGCGCACCTTTGCCCTGGAGCGGA 3451

     124 ..............ArgTyrThrIleProValLeuHisValLysAsnTrp 135
                       |||:::   :::     :::|||   ::::::|||
    3452 GAGGTTACTCAGCCCGGCATGAGGTCCGCCAGTTCCATTTCACAGCGTGG 3501

     136 ProAspHis...GlnProLeuProSerThrAspGlnLeuGluTyrLeuAl 151
         |||:::|||     |||      :::|||:::   |||:::::::::
    3502 CCAGAGCATGGTGTCCCCTACCACGCCACGGGGCTGCTGGCCTTCATC.. 3549

     151 aAspArgValLysAsnSerAsnGlnAsnGlyAlaProGlyArgSerSerS 168
           |||||||||:::|||:::::   :::|||
    3550 .CGGCGTGTGAAGGCTTCCACTCCACCTGATGCCGGGCCCATT....... 3591

     168 erAspLysHisLeuProMetIleHisCysLeuGlyGlyValGlyArgThr 184
                       :::|||||||||   :::|||:::||||||||||
    3592 .................GTCATTCACTGCAGTGCAGGAACTGGCCGCACA 3624

     185 Gly 185
         |||
    3625 GGC 3627

seq_name: /cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-991-258A-1

seq_documentation_block:

; Sequence 1, Application US/08991258A
; Patent No. 5928887
;  GENERAL INFORMATION:
;    APPLICANT:  Cheng, Jill
;    APPLICANT:  Lasky, Laurence A.
;    TITLE OF INVENTION:  A NOVEL KAPPA/MU-LIKE PROTEIN TYROSINE
;    TITLE OF INVENTION:  PHOSPHATASE, PTP LAMBDA
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT, LLP
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/991,258A
;      FILING DATE:  17-DEC-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/652,971
;      FILING DATE:  24-MAY-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dreger, Walter H.
;      REGISTRATION NUMBER:  24,190
;      REFERENCE/DOCKET NUMBER:  A-63478-3/WHD/MTK
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5769 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  379..4686
US-08-991-258A-1


alignment_scores:
            Quality:   89.00              Length:    151
              Ratio:  1.011                Gaps:      7
 Percent Similarity:  58.278   Percent Identity: 23.179

alignment_block:
 US-09-068-804-12 x US-08-991-258A-1   ..

 Align seg 1/1  to: US-08-991-258A-1  from: 1  to: 5769

      43 IleArgThrGlnLysIleArgProAspArgLeuGluAlaHisMetLysMe 59
         |||   ||||||      :::|||:::   :::   :::      :::|| 
    3223 ATAGCCACTCAAGGGCCAAAGCCTGAGATGATCTACGATTTCTGGCGCAT 3272

      59 tLeuLeuGluLysGluCysSerCysLeuValValLeuThrSerGluArgS 76
         |:::   :::   :::|||:::::   :::|||:::::::::
    3273 GGTGTGGCAGGAACAGTGTGCGAGC...ATCGTCATGATCACCAAGCTGG 3319

      76 erAspAlaGlyLysThrIleThrThrTyrPheArg......GlySerTyr 90
           ::::::|||:::::      :::   :::::      :::|||
    3320 TAGAGGTGGGCAGGGTGAAGTGTTCTCGCTACTGGCCTGAGGACTCAGAC 3369

      91 ThrPheGlyGluValHisThrAsnSerGlnLysValSerSerAlaSerGl 107
           :::|||:::::   :::::      |||:::
    3370 ATGTATGGGGACATCAAGATCACGCTGGTAAAGACA.............. 3405
```

```
    107 nGlyGluAlaIleAspGlnTyrAsnMetGlnLeuSerCysGlyGluLys. 123
    3406 ....GAGACACTGGCTGAGTATGTGGTGCGCACCTTTGCCCTGGAGCGGA 3451

     124 .............ArgTyrThrIleProValLeuHisValLysAsnTrp 135
    3452 GAGGTTACTCAGCCCGGCATGAGGTCCGCCAGTTCCATTTCACAGCGTGG 3501

     136 ProAspHis...GlnProLeuProSerThrAspGlnLeuGluTyrLeuAl 151
    3502 CCAGAGCATGGTGTCCCCTACCACGCCACGGGGCTGCTGGCCTTCATC.. 3549

     151 aAspArgValLysAsnSerAsnGlnAsnGlyAlaProGlyArgSerSerS 168
    3550 .CGGCGTGTGAAGGCTTCCACTCCACCTGATGCCGGGCCCATT..... 3591

     168 erAspLysHisLeuProMetIleHisCysLeuGlyGlyValGlyArgThr 184
    3592 .................GTCATTCACTGCAGTGCAGGAACTGGCCGCACA 3624

     185 Gly 185
    3625 GGC 3627

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-202-389-11

seq_documentation_block:
; Sequence 11, Application US/08202389
; Patent No. 5536636
;  GENERAL INFORMATION:
;    APPLICANT:  Freeman Jr., Robert M.
;    APPLICANT:  Plutzky, Jorge
;    APPLICANT:  Neel, Benjamin G.
;    APPLICANT:  Rosenberg, Robert D.
;    TITLE OF INVENTION:  IDENTIFICATION OF NOVEL TYROSINE
;    TITLE OF INVENTION:  PHOSPHATASES HAVING SH2 DOMAINS
;    NUMBER OF SEQUENCES:  54
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
;      STREET:  Two Militia Drive
;      CITY:  Lexington
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/202,389
;      FILING DATE:  28-FEB-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/983,926
;      FILING DATE:  01-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/829,141
;      FILING DATE:  31-JAN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/721,112
;      FILING DATE:  26-JUN-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Granahan, Patricia
;      REGISTRATION NUMBER:  32,227
;      REFERENCE/DOCKET NUMBER:  BIH92-05MA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 861-6240
;      TELEFAX:  (617) 861-9540
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2276 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  114..1893
US-08-202-389-11


alignment_scores:
           Quality:   84.50            Length:    242
             Ratio:  0.710              Gaps:      9
 Percent Similarity:  49.174   Percent Identity: 20.248

alignment_block:
 US-09-068-804-12 x US-08-202-389-11   ..

 Align seg 1/1  to: US-08-202-389-11  from: 1  to: 2276

      14 GlnGlnThrGlnValLysLeuSerAspGlyMetPro..... 25
     969 GATCATACCAGGGTTGTCCTACACGATGGTGATCCCAATGAGCCTGTTTC 1018

      26 .............ValProValIleArgLeuThrPheAspGlyLysProV 38
    1019 AGATTACATCAATGCAAATATCATCATGCCTGAATTTGAAACCAAGTGCA 1068

      38 alAlaLeuAlaGly.............IleArgThrGlnLysIleArgPro 50
    1069 ACAATTCAAAGCCCAAAAAGAGTTACATTGCCACACAAGGCTGCCTGCAA 1118

      51 AspArgLeuGluAlaHisMetLysMetLeuLeuGluLysGluCysSerCy 67
    1119 AACACGGTGAATGACTTTTGGCGGATGGTGTTCCAAGAAAACTCCCGAGT 1168

      67 sLeuValValLeuThrSerGluArgSerAspAlaGlyLysThrIleThrT 84
    1169 GATTGTCATGACAACGAAAGAAGTGGAGAGAGGAAAGAGTAAATGTGTCA 1218

      84 hrTyrPheArgGlySerTyrThr.........PheGlyGluValHisThr 97
    1219 AATACTGGCCTGATGAGTATGCTCTAAAAGAATATGGCGTCATGCGTGTT 1268

      98 AsnSerGlnLysValSerSerAlaSerGlnGlyGluAlaIleAspGlnTy 114
    1269 AGGAACGTCAAAGAAAGCGCCGCT..................CATGACTA 1300

     114 rAsnMetGln..................LeuSerCysGlyGluLysArgT 125
    1301 TACGCTAAGAGAACTTAAACTTTCAAAGGTTGGACAAGGGAATACGGAGA 1350

     125 yrThrIleProValLeuHisValLysAsnTrpProAspHis...GlnPro 140
    1351 GAACGGTCTGGCAATACCACTTTCGGACCTGGCCGGACCACGGCGTGCCC 1400

     141 LeuProSerThrAspGlnLeuGluTyrLeuAlaAsp......ArgValLy 155
    1401 AGCGACCCTGGGGGCGTGCTGGACTTCCTGGAGGAGGTGCACCATAAGCA 1450

     155 sAsnSerAsnGlnAsnGlyAlaProGlyArgSerSerSerAspLysHisL 172
    1451 GGAGAGCATCATGGATGCAGGGCCG..................... 1475

     172 euProMetIleHisCysLeuGlyGlyValGlyArgThrGlyThrMetAla 188
    1476 ..GTCGTGGTGCACTGCAGTGCTGGAATTGGCCGGACAGGGACGTTCATT 1523

     189 AlaAlaLeuValLeuLysAsp..... 195
    1524 GTGATTGATATTCTTATTGACATCATCAGAGAGAAAGGTGTTGACTGCGA 1573

     196 .............................AsnProHisSerAsnL 201
```

```
                                                    :::::::::|||::::
    1574 TATTGACGTTCCCAAAACCATCCAGATGGTGCGGTCTCAGAGGTCAGGGA 1623

     201 euGluGlnValArgAlaAspPheArg 209
         ::   |||:::   |||:::::|||
    1624 TGGTCCAGACAGAAGCACAGTACCGA 1649

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-018-129-4

seq_documentation_block:
; Sequence 4, Application US/08018129
; Patent No. 5589375
;  GENERAL INFORMATION:
;    APPLICANT:  Ullrich, Axel
;    APPLICANT:  Vogel, Wolfgang
;    TITLE OF INVENTION:  PTP 1D: A NOVEL PROTEIN TYROSINE
;    TITLE OF INVENTION:  PHOSPHATASE
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PENNIE & EDMONDS
;      STREET:  1155 Avenue of Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/018,129
;      FILING DATE:  19930216
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Misrock, S. Leslie
;      REGISTRATION NUMBER:  18,872
;      REFERENCE/DOCKET NUMBER:  7683-017
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-8864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2790 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  130..1911
US-08-018-129-4


alignment_scores:
            Quality:   84.00              Length:    243
              Ratio:  0.694                Gaps:      9
 Percent Similarity:  49.794   Percent Identity: 19.753

alignment_block:
 US-09-068-804-12 x US-08-018-129-4    ..

 Align seg 1/1  to: US-08-018-129-4  from: 1  to: 2790

       4 ThrIleCysGlyGlnGlnHisSerIleAsnGlnGlnThrGlnValLysLe 20
         |||||||||:::::::::::::::::::
    1006 ACGATCTGTGATCCCAATGAGCCTGTT..................... 1032

      20 uSerAspGlyMetProValProValIleArgLeuThrPheAspGlyLysP 37
         ||||||   :::   :::   :::|||         |||:::   |||
    1033 .TCAGATTACATCAATGCAAATATCATCATGCCTGAATTTGAAACCAAGT 1081

      37 roValAlaLeuAlaGly...........IleArgThrGlnLysIleArg 49
         :::                        |||   ||||||
    1082 GCAACAATTCAAAGCCCAAAAAGAGTTACATTGCCACACAAGGCTGCCTG 1131

      50 ProAspArgLeuGluAlaHisMetLysMetLeuLeuGluLysGluCysSe 66
         ::::::   :::::::::      :::|||:::::::::   ::::::
    1132 CAAAACACGGTGAATGACTTTTGGCGGATGGTGTTCCAAGAAAACTCCCG 1181

      66 rCysLeuValValLeuThrSerGluArgSerAspAlaGlyLysThrIleT 83
         :::|||:::   |||   |||         :::   :::   :
    1182 AGTGATTGTCATGACAACGAAAGAAGTGGAGAGAGGAAAGAGTAAATGTG 1231

      83 hrThrTyrPheArgGlySerTyrThr.........PheGlyGluValHis 96
         ::   |||      :::   |||:::         :::|||   ::::::
    1232 CTCAATACTGGCCTGATGAGTATGCTCTAAAAGAATATGGCGTCATGCGT 1281

      97 ThrAsnSerGlnLysValSerSerAlaSerGlnGlyGluAlaIleAspGl 113
         :::   :::   |||   |||:::|||                   :::::
    1282 GTTAGGAACGTCAAAGAAAGCGCCGCT.................CATGA 1313

     113 nTyrAsnMetGln.................LeuSerCysGlyGluLysA 124
         :|||::::::::::                 ::::::   |||::::::
    1314 CTATACGCTAAGAGAACTTAAACTTTCAAAGGTTGGACAAGGGAATACGG 1363

     124 rgTyrThrIleProValLeuHisValLysAsnTrpProAspHis...Gln 139
         |||:::      |||   ::::::||||||||||||
    1364 AGAGAACGGTCTGGCAATACCACTTTCGGACCTGGCCGGACCACGGCGTG 1413

     140 ProLeuProSerThrAspGlnLeuGluTyrLeuAlaAsp......ArgVa 154
         |||      :::   :::   |||::::::|||::::::      :::
    1414 CCCAGCGACCCTGGGGGCGTGCTGGACTTCCTGGAGGAGGTGCACCATAA 1463

     154 lLysAsnSerAsnGlnAsnGlyAlaProGlyArgSerSerSerAspLysH 171
         :::::::|||      :::::::::|||
    1464 GCAGGAGAGCATCATGGATGCAGGGCCG..................... 1491

     171 isLeuProMetIleHisCysLeuGlyGlyValGlyArgThrGlyThrMet 187
             ::::::||||||   :::|||:::||||||||||||||||
    1492 .....GTCGTGGTGCACTGCAGTGCTGGAATTGGCCGGACAGGGACGTTC 1536

     188 AlaAlaAlaLeuValLeuLysAsp........................... 195
         :::      :::|||   |||
    1537 ATTGTGATTGATATTCTTATTGACATCATCAGAGAGAAAGGTGTTGACTG 1586

     196 .....................................AsnProHisSerA 200
                                             :::::::::|||:
    1587 CGATATTGACGTTCCCAAAACCATCCAGATGGTGCGGTCTCAGAGGTCAG 1636

     200 snLeuGluGlnValArgAlaAspPheArg 209
         :::::   |||:::   |||::::::|||
    1637 GGATGGTCCAGACAGAAGCACAGTACCGA 1665

seq_name: /cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-791-522-2

seq_documentation_block:
; Sequence 2, Application US/08791522
; Patent No. 5935817
;  GENERAL INFORMATION:
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Goli, Surya K.
;    TITLE OF INVENTION:  NOVEL HUMAN CYSTATIN-LIKE
;    TITLE OF INVENTION:  PROTEIN
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
```

```
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/791,522
;      FILING DATE:  Filed Herewith
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0193 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  924 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  30443
US-08-791-522-2


alignment_scores:
            Quality:   79.00              Length:     95
              Ratio:  1.491                Gaps:      5
 Percent Similarity:  55.789   Percent Identity: 27.368

alignment_block:
 US-09-068-804-12 x US-08-791-522-2   ..

 Align seg 1/1  to: US-08-791-522-2  from: 1  to: 924

     128 ProValLeuHisValLysAsnTrpProAspHisGlnProLeuProSerTh 144
         |||||||||       |||:::|||   :::|||:::|||||||||:::::
      87 CCCGTGCTGCCTGAGAAGGATTGGCACGGGCACAGACCACTGCCCCCACC 136

     144 rAspGlnLeuGluTyrLeuAlaAspArgValLysAsnSerAsnGlnAsnG 161
         ::::   |||   :::|||:::   :::::::::::::::::::|||
     137 TGCC...CTGCGCCATCTACCCAAGAAGGCTCGGCACGGGCACCAACCAC 183

     161 lyAlaProGlyArgSerSerSerAspLysHisLeuProMetIleHisCys 177
           :::|||      :::::::::   ::::::            ||||||
     184 TGCCTCCAACTGCCCCATGCTGCCTGAGAAGG............CACTGC 221

     178 LeuGlyGlyValGlyArgThrGlyThrMetAla...AlaAlaLeuValLe 193
                       |||:::|||   :::   :::||||||   ||
     222 ..................ACGGCCACCCCCAACTGCCCCGCACTGTCCCT 253

     193 uLysAspAsnProHisSerAsnLeuGluGlnValArgAlaAspPheArgI 210
         |   ::::::|||       ||||||
     254 ACCCGGGCAGCCATGCGAGCGGCTGGAACTC..................... 284

     210 leHisGlyThrIleGluCysTrpLysThrProAla 221
                         ||||||   :::::::|||
     285 .................TGCTGGCCTTCTGCTGCC 302

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-202-389-7

seq_documentation_block:
; Sequence 7, Application US/08202389
; Patent No. 5536636
;  GENERAL INFORMATION:
;    APPLICANT:  Freeman Jr., Robert M.
;    APPLICANT:  Plutzky, Jorge
;    APPLICANT:  Neel, Benjamin G.
;    APPLICANT:  Rosenberg, Robert D.
;    TITLE OF INVENTION:  IDENTIFICATION OF NOVEL TYROSINE
;    TITLE OF INVENTION:  PHOSPHATASES HAVING SH2 DOMAINS
;    NUMBER OF SEQUENCES:  54
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
;      STREET:  Two Militia Drive
;      CITY:  Lexington
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/202,389
;      FILING DATE:  28-FEB-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/983,926
;      FILING DATE:  01-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/829,141
;      FILING DATE:  31-JAN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/721,112
;      FILING DATE:  26-JUN-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Granahan, Patricia
;      REGISTRATION NUMBER:  32,227
;      REFERENCE/DOCKET NUMBER:  BIH92-05MA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 861-6240
;      TELEFAX:  (617) 861-9540
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2143 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  145..2037
US-08-202-389-7


alignment_scores:
            Quality:   82.50              Length:    167
              Ratio:  1.058                Gaps:     10
 Percent Similarity:  46.707   Percent Identity: 22.156

alignment_block:
 US-09-068-804-12 x US-08-202-389-7   ..

 Align seg 1/1  to: US-08-202-389-7  from: 1  to: 2143

     119 SerCysGlyGlu.........LysArgTyrThrIleProValLeuHisVa 132
         :::|||||||||         :::   |||:::::   :::|||:::||
    1285 AACTGCGGGGAGCATGACACAACCGAATACAAACTCCGTACCTTACAGGT 1334

     132 lLys..................................................... 133
         |
    1335 CTCCCCGCTGGACAATGGAGACCTGATTCGGGAGATCTGGCATTACCAGT 1384

     134 .....AsnTrpProAspHis...GlnProLeuProSerThrAspGlnLeu 147
              :::|||||||||||||      |||      :::   :::   |||
    1385 ACCTGAGCTGGCCCGACCATGGGGTCCCCAGTGAGCCTGGGGGTGTCCTC 1434

     148 GluTyrLeuAlaAspArgValLysAsnSerAsnGlnAsnGlyAlaProGl 164
           :::|||   |||:::::::::::   :::::::::
    1435 AGCTTCCTG...GACCAGATCAACCAGCGGCAGGAAAGT.......... 1470

     164 yArgSerSerSerAspLysHisLeuPro.............MetIleH 176
```

```
                              ||||||                  :::::::|
    1471 .....................CTGCCTCACGCAGGGCCCATCATCGTGC 1498

     176 isCysLeuGlyGlyValGlyArgThrGlyThrMetAlaAlaAlaLeuVal 192
         |||||   :::|||:::||||||||||||||||:::   :::      :::
    1499 ACTGCAGCGCCGGCATCGGCCGCACAGGCACCATCATTGTCATCGACATG 1548

     193 LeuLysAspAsn..................................... 196
         |||:::::|||
    1549 CTCATGGAGAACATCTCCACCAAGGGCCTGGACTGTGACATTGACATCCA 1598

     197 .........................ProHisSerAsnLeuGluGlnValA 205
                                 :::::::|||::::::   |||:::
    1599 GAAGACCATCCAGATGGTGCGGGCGCAGCGCTCGGGCATGGTGCAGACGG 1648

     205 rgAlaAspPheArg...IleHisGlyThrIleGluCysTrpLysThrPro 220
           |||:::::::::   |||:::   :::|||
    1649 AGGCGCAGTACAAGTTCATCTACGTGGCCATC..................... 1680

     221 AlaGlnPheValGln..............LeuLysAlaMetGlnAlaGl 232
         |||||||||:::::                |||   :::::::|||:::||
    1681 GCCCAGTTCATTGAAACCACTAAGAAGAAGCTGGAGGTCCTGCAGTCGCA 1730

     232 n 232
         |
    1731 G 1731
```

seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-201-697-3

```
seq_documentation_block:
; Sequence 3, Application US/08201697
; Patent No. 5705623
;  GENERAL INFORMATION:
;    APPLICANT:  Wiggins, Roger C.
;    APPLICANT:  Thomas, Peedikayil E.
;    TITLE OF INVENTION:  Mammalian Glomerular Epithelial Protein
;    TITLE OF INVENTION:  1
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell and Flores
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/201,697
;      FILING DATE:  25-FEB-1994
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryb A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-UM 9783
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4752 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  808..4372
US-08-201-697-3
```

```
alignment_scores:
            Quality:   84.50              Length:    198
              Ratio:  0.775                Gaps:     10
 Percent Similarity:  55.051   Percent Identity: 20.202

alignment_block:
 US-09-068-804-12 x US-08-201-697-3   ..

 Align seg 1/1  to: US-08-201-697-3  from: 1  to: 4752

      43 IleArgThrGlnLysIleArgProAspArgLeuGluAlaHisMetLysMe 59
         |||   ||||||          |||:::      ::::::      |||||
    3745 ATTGCCACCCAGGGGCCACTGCCTGAAACCAGAAATGACTTCTGGAAGAT 3794

      59 tLeuLeuGluLysGluCysSerCysLeuValValLeuThrSerGluArgS 76
         |:::|||::::::   :::      :::|||:::||||||      ::::
    3795 GGTCCTGCAACAAAAGTCTCAGATTATTGTCATGCTCACT......CAGT 3838

      76 erAspAlaGlyLysThrIleThr..............ThrTyrPheArg 87
         ::::::::   :::   ::::::                ::::::
    3839 GTAATGAGAAAAGGAGGGTGAAATGTGACCATTACTGGCCATTCACGGAA 3888

      88 GlySerTyrThrPheGlyGluValHisThrAsnSerGlnLysValSerSe 104
         :::   ::::::|||:::::      :::   ::::::::|||
    3889 GAACCTATAGCCTATGGAGACATC......ACTGTGGAGATGATTTCAGA 3932

     104 rAlaSerGlnGlyGlu...AlaIleAspGlnTyrAsnMetGlnLeuSerC 120
          :::   |||::::::   |||      ::::::      ::::::::::
    3933 GGAAGAGCAGGACGACTGGGCCTGTAGACACTTC......CGGATCAACT 3976

     120 ysGlyGluLysArgTyrThrIleProValLeuHisValLysAsnTrpPro 136
         ::::::::            :::      ::::::   ::::::||||||
    3977 ATGCTGACGAGATGCAGGATGTGATGCATTTTAACTACACTGCATGGCCT 4026

     137 AspHisGlnProLeuProSerThrAspGlnLeuGluTyrLeuAlaAspAr 153
         ||||||      :::|||:::::::      |||   :::   :::
    4027 GATCAT...GGTGTGCCCACAGCAAATGCTGCAGAAAGTATCCTGCAGTT 4073

     153 gValLysAsnSerAsnGlnAsnGlyAlaProGlyArgSerSerSerAspL 170
          |||            |||::::::::::   :::::::::
    4074 TGTACACATGGTCCGACAGCAAGCTACCAAGAGCAAAGGT.......... 4113

     170 ysHisLeuProMet...IleHisCysLeuGlyGlyValGlyArgThrGly 185
                |||||   |||||||||   :::|||||||||||||||||||
    4114 ........CCCATGATCATTCACTGCAGTGCTGGCGTGGGACGGACAGGA 4155

     186 ThrMetAlaAlaAlaLeuValLeuLysAspAsnProHisSerAsnLeuGl 202
         |||      |||                                   |||::
    4156 ACATTCATTGCC................................CTGGA 4172

     202 uGlnValArgAlaAspPheArgIleHisGlyThrIleGluCysTrpLysT 219
         :::::::      :::::::|||   |||
    4173 CAGGCTCTTGCAGCACATTCGGGATCAT..................... 4200

     219 hrProAlaGlnPheValGlnLeuLysAlaMetGlnAlaGlnLeu 233
                :::|||||||::::::   ::::::   :::::::::
    4201 ........GAGTTTGTTGACATCTTAGGGCTGGTGTCAGAAATG 4236
```

seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-201-697-6

```
seq_documentation_block:
; Sequence 6, Application US/08201697
; Patent No. 5705623
;  GENERAL INFORMATION:
;    APPLICANT:  Wiggins, Roger C.
;    APPLICANT:  Thomas, Peedikayil E.
;    TITLE OF INVENTION:  Mammalian Glomerular Epithelial Protein
;    TITLE OF INVENTION:  1
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell and Flores
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
```

```
;       STATE:  California
;       COUNTRY:  USA
;       ZIP:  92122
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/201,697
;       FILING DATE:  25-FEB-1994
;       CLASSIFICATION:  530
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Campbell, Cathryb A.
;       REGISTRATION NUMBER:  31,815
;       REFERENCE/DOCKET NUMBER:  P-UM 9783
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (619) 535-9001
;       TELEFAX:  (619) 535-8949
;   INFORMATION FOR SEQ ID NO:  6:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  4799 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
US-08-201-697-6


alignment_scores:
            Quality:   84.50              Length:    198
              Ratio:  0.775                Gaps:     10
 Percent Similarity:  55.051   Percent Identity: 20.202

alignment_block:
 US-09-068-804-12 x US-08-201-697-6   ..

 Align seg 1/1  to: US-08-201-697-6  from: 1  to: 4799

      43 IleArgThrGlnLysIleArgProAspArgLeuGluAlaHisMetLysMe 59
         |||   ||||||         |||:::      ::::::      |||||
    3792 ATTGCCACCCAGGGGCCACTGCCTGAAACCAGAAATGACTTCTGGAAGAT 3841

      59 tLeuLeuGluLysGluCysSerCysLeuValValLeuThrSerGluArgS 76
         |:::|||::::::   :::      :::|||:::||||||      ::::
    3842 GGTCCTGCAACAAAAGTCTCAGATTATTGTCATGCTCACT......CAGT 3885

      76 erAspAlaGlyLysThrIleThr..............ThrTyrPheArg 87
         ::::::::   :::   ::::::               ::::::
    3886 GTAATGAGAAAAGGAGGGTGAAATGTGACCATTACTGGCCATTCACGGAA 3935

      88 GlySerTyrThrPheGlyGluValHisThrAsnSerGlnLysValSerSe 104
            :::   ::::::|||:::::      :::   ::::::::::|||
    3936 GAACCTATAGCCTATGGAGACATC......ACTGTGGAGATGATTTCAGA 3979

     104 rAlaSerGlnGlyGlu...AlaIleAspGlnTyrAsnMetGlnLeuSerC 120
          :::   |||:::::   |||      ::::::      ::::::::::
    3980 GGAAGAGCAGGACGACTGGGCCTGTAGACACTTC......CGGATCAACT 4023

     120 ysGlyGluLysArgTyrThrIleProValLeuHisValLysAsnTrpPro 136
         ::::::::              :::      ::::::   ::::::||||||
    4024 ATGCTGACGAGATGCAGGATGTGATGCATTTTAACTACACTGCATGGCCT 4073

     137 AspHisGlnProLeuProSerThrAspGlnLeuGluTyrLeuAlaAspAr 153
         ||||||      :::|||::::::::::      |||   :::   :::
    4074 GATCAT...GGTGTGCCCACAGCAAATGCTGCAGAAAGTATCCTGCAGTT 4120

     153 gValLysAsnSerAsnGlnAsnGlyAlaProGlyArgSerSerSerAspL 170
          |||           |||::::::::   :::::::::
    4121 TGTACACATGGTCCGACAGCAAGCTACCAAGAGCAAAGGT.......... 4160

     170 ysHisLeuProMet...IleHisCysLeuGlyGlyValGlyArgThrGly 185
             ||||||   |||||||||   :::|||||||||||||||||||
    4161 T........CCCATGATCATTCACTGCAGTGCTGGCGTGGGACGGACAGGA 4202

     186 ThrMetAlaAlaAlaLeuValLeuLysAspAsnProHisSerAsnLeuGl 202
         |||      |||                                  |||::
    4203 ACATTCATTGCC.............................CTGGA 4219

     202 uGlnValArgAlaAspPheArgIleHisGlyThrIleGluCysTrpLysT 219
         :::::::      ::::::|||   |||
    4220 CAGGCTCTTGCAGCACATTCGGGATCAT..................... 4247

     219 hrProAlaGlnPheValGlnLeuLysAlaMetGlnAlaGlnLeu 233
             :::||||||::::::   ::::::   :::::::::
    4248 ........GAGTTTGTTGACATCTTAGGGCTGGTGTCAGAAATG 4283

seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-201-697-5

seq_documentation_block:
; Sequence 5, Application US/08201697
; Patent No. 5705623
;  GENERAL INFORMATION:
;    APPLICANT:  Wiggins, Roger C.
;    APPLICANT:  Thomas, Peedikayil E.
;    TITLE OF INVENTION:  Mammalian Glomerular Epithelial Protein
;    TITLE OF INVENTION:  1
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell and Flores
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/201,697
;      FILING DATE:  25-FEB-1994
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryb A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-UM 9783
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4815 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-201-697-5


alignment_scores:
            Quality:   84.50              Length:    198
              Ratio:  0.775                Gaps:     10
 Percent Similarity:  55.051   Percent Identity: 20.202

alignment_block:
 US-09-068-804-12 x US-08-201-697-5   ..

 Align seg 1/1  to: US-08-201-697-5  from: 1  to: 4815

      43 IleArgThrGlnLysIleArgProAspArgLeuGluAlaHisMetLysMe 59
         |||   ||||||         |||:::      ::::::      |||||
    3808 ATTGCCACCCAGGGGCCACTGCCTGAAACCAGAAATGACTTCTGGAAGAT 3857

      59 tLeuLeuGluLysGluCysSerCysLeuValValLeuThrSerGluArgS 76
         |:::|||::::::   :::      :::|||:::||||||      ::::
    3858 GGTCCTGCAACAAAAGTCTCAGATTATTGTCATGCTCACT......CAGT 3901
```

```
     76 erAspAlaGlyLysThrIleThr.............ThrTyrPheArg 87
        ::::::::   :::   ::::::                ::::::
   3902 GTAATGAGAAAAGGAGGGTGAAATGTGACCATTACTGGCCATTCACGGAA 3951

     88 GlySerTyrThrPheGlyGluValHisThrAsnSerGlnLysValSerSe 104
         :::   ::::::|||:::::::      :::   ::::::::::|||
   3952 GAACCTATAGCCTATGGAGACATC......ACTGTGGAGATGATTTCAGA 3995

    104 rAlaSerGlnGlyGlu...AlaIleAspGlnTyrAsnMetGlnLeuSerC 120
         :::   |||::::::   |||      ::::::      ::::::::::
   3996 GGAAGAGCAGGACGACTGGGCCTGTAGACACTTC......CGGATCAACT 4039

    120 ysGlyGluLysArgTyrThrIleProValLeuHisValLysAsnTrpPro 136
        ::::::::            :::      ::::::   ::::::||||||
   4040 ATGCTGACGAGATGCAGGATGTGATGCATTTTAACTACACTGCATGGCCT 4089

    137 AspHisGlnProLeuProSerThrAspGlnLeuGluTyrLeuAlaAspAr 153
        ||||||      :::|||::::::::::      |||   :::   :::
   4090 GATCAT...GGTGTGCCCACAGCAAATGCTGCAGAAAGTATCCTGCAGTT 4136

    153 gValLysAsnSerAsnGlnAsnGlyAlaProGlyArgSerSerSerAspL 170
         |||           |||:::::::::   :::::::::
   4137 TGTACACATGGTCCGACAGCAAGCTACCAAGAGCAAAGGT.......... 4176

    170 ysHisLeuProMet...IleHisCysLeuGlyGlyValGlyArgThrGly 185
             ||||||   |||||||||   :::||||||||||||||||||
   4177 ........CCCATGATCATTCACTGCAGTGCTGGCGTGGGACGGACAGGA 4218

    186 ThrMetAlaAlaAlaLeuValLeuLysAspAsnProHisSerAsnLeuGl 202
        |||     |||                                  |||::
   4219 ACATTCATTGCC................................CTGGA 4235

    202 uGlnValArgAlaAspPheArgIleHisGlyThrIleGluCysTrpLysT 219
        :::::::      :::::::|||   |||
   4236 CAGGCTCTTGCAGCACATTCGGGATCAT..................... 4263

    219 hrProAlaGlnPheValGlnLeuLysAlaMetGlnAlaGlnLeu 233
             :::||||||:::::::   ::::::   :::::::::
   4264 ........GAGTTTGTTGACATCTTAGGGCTGGTGTCAGAAATG 4299

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-015-985-4

seq_documentation_block:
; Sequence 4, Application US/08015985
; Patent No. 5538886
;  GENERAL INFORMATION:
;    APPLICANT:  Schlessinger, Joseph
;    APPLICANT:  Sap, Jan M.
;    TITLE OF INVENTION:  NOVEL RECEPTOR-TYPE PHOSPHOTYROSINE
;    TITLE OF INVENTION:  PHOSPHATASE-ALPHA
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PENNIE & EDMONDS
;      STREET:  1155 AVENUE OF THE AMERICAS
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/015,985
;      FILING DATE:  10-FEB-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Coruzzi, Laura A.
;      REGISTRATION NUMBER:  30,742
;      REFERENCE/DOCKET NUMBER:  7683-020
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-9741/8864
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2872 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
US-08-015-985-4


alignment_scores:
            Quality:   81.50              Length:    174
              Ratio:  0.832                Gaps:      8
 Percent Similarity:  56.322   Percent Identity: 22.989

alignment_block:
 US-09-068-804-12 x US-08-015-985-4   ..

 Align seg 1/1  to: US-08-015-985-4  from: 1  to: 2872

      43 IleArgThrGlnLysIleArgProAspArgLeuGluAlaHisMetLysMe 59
         |||   :::|||      :::   :::   :::::::::      :::||
    1162 ATCGCTGCACAAGGACCAAAAGAAGAAACAGTGAATGACTTCTGGAGAAT 1211

      59 tLeuLeuGluLysGluCysSerCysLeuValValLeuThrSer......G 74
         |:::   |||::::::   :::   :::|||:::::|||:::      |
    1212 GATATGGGAACAAAACACAGCTACTATTGTCATGGTGACCAACCTGAAGG 1261

      74 luArgSerAspAlaGlyLysThrIleThrThrTyrPhe......ArgGly 88
         |||||   :::          :::   :::   |||          :::|||
    1262 AGAGAAAGGAGTGT......AAATGTGCCCAATACTGGCCAGACCAAGGC 1305

      89 SerTyrThrPheGlyGluValHisThrAsnSerGlnLysValSerSerAl 105
         :::   |||:::|||:::|||:::::::::   :::   |||:::
    1306 TGCTGGACCTATGGGAATGTCCGTGTGTCTGTCGAGGATGTGACTGTTCT 1355

     105 aSerGln..............................GlyGluAlaI 111
            :::                                |||:::::::
    1356 GGTGGACTACACAGTACGGAAATTCTCGATCCAGCAGGTGGGCGACGTGA 1405

     111 leAspGlnTyrAsnMetGlnLeuSerCysGlyGluLysArgTyrThrIle 127
         ::::::::          :::|||                      |||
    1406 CCAACAGGAAACCACAGCGCCTC.......................ATC 1431

     128 ProValLeuHisValLysAsnTrpProAspHis...GlnProLeuProSe 143
         :::   :::|||   ::::::|||||||||         |||:::::::::
    1432 ACTCAGTTCCACTTCACCAGCTGGCCAGACTTTGGGGTGCCTTTCACCCC 1481

     143 rThrAspGlnLeuGluTyrLeuAlaAspArgValLysAsnSerAsnGlnA 160
         :::::::   |||   :::|||      :::||||||::::::|||::::
    1482 AATTGGCATGCTCAAGTTCCTC...AAGAAGGTGAAGGCCTGTAACCCTC 1528

     160 snGlyAlaProGlyArgSerSerSerAspLysHisLeuProMetIleHis 176
         ::   |||   :::                          ::::::|||
    1529 AGTACGCAGGGGCTATC.......................GTGGTCCAC 1554

     177 CysLeuGlyGlyValGlyArgThrGlyThrMetAlaAlaAlaLeuValLe 193
         |||   :::|||||||||||||||||||||   ::::::      :::::
    1555 TGCAGTGCAGGTGTAGGGCGCACTGGCACCTTTGTTGTCATCGATGCCAT 1604

     193 uLysAspAsnProHisSerAsn 200
         :   |||      ||||||:::
    1605 GCTGGACATGATGCATTCGGAG 1626

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-015-985-2

seq_documentation_block:
; Sequence 2, Application US/08015985
; Patent No. 5538886
;  GENERAL INFORMATION:
```

```
;    APPLICANT:  Schlessinger, Joseph
;    APPLICANT:  Sap, Jan M.
;    TITLE OF INVENTION:  NOVEL RECEPTOR-TYPE PHOSPHOTYROSINE
;    TITLE OF INVENTION:  PHOSPHATASE-ALPHA
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PENNIE & EDMONDS
;      STREET:  1155 AVENUE OF THE AMERICAS
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/015,985
;      FILING DATE:  10-FEB-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Coruzzi, Laura A.
;      REGISTRATION NUMBER:  30,742
;      REFERENCE/DOCKET NUMBER:  7683-020
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-9741/8864
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2409 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
US-08-015-985-2


alignment_scores:
            Quality:   80.50            Length:    158
              Ratio:  0.875              Gaps:      8
 Percent Similarity:  58.228   Percent Identity: 22.785

alignment_block:
 US-09-068-804-12 x US-08-015-985-2   ..

 Align seg 1/1  to: US-08-015-985-2  from: 1  to: 2409

      43 IleArgThrGlnLysIleArgProAspArgLeuGluAlaHisMetLysMe 59
         |||   :::|||      :::   :::   :::::::::      :::||
     931 ATTGCTGCACAAGGACCAAAAGAAGAAACGGTGAATGATTTCTGGCGGAT 980

      59 tLeuLeuGluLysGluCysSerCysLeuValValLeuThrSer......G 74
         |:::   |||::::::   :::   :::|||::::::|||:::      |
     981 GATCTGGGAACAAAACACAGCCACCATCGTCATGGTTACCAACCTGAAGG 1030

      74 luArgSerAspAlaGlyLysThrIleThrThrTyrPhe......ArgGly 88
         |||||   :::         :::   :::   |||          :::|||
    1031 AGAGAAAGGAGTGC......AAGTGCGCCCAGTACTGGCCAGACCAAGGC 1074

      89 SerTyrThrPheGlyGluValHisThrAsnSerGlnLysValSerSerAl 105
         :::   |||:::|||:::::::::::::::::   :::   |||:::
    1075 TGCTGGACCTATGGGAATATTCGGGTGTCTGTAGAGGATGTGACT..... 1119

     105 aSerGlnGlyGluAlaIleAspGlnTyrAsnMetGlnLeuSerCysGlyG 122
                   ::::::   :::|||:::::::::      |||   :
    1120 .............GTCCTGGTGGACTACACAGTACGGAAGTTCTGCATCC 1156

     122 luLys.........................ArgTyrThrIleProVal 129
         :::::                       :::      |||:::
    1157 AGCAGGTGGGCGACATGACCAACAGAAAGCCACAGCGCCTCATCACTCAG 1206

      130 LeuHisValLysAsnTrpProAspHis...GlnProLeuProSerThrAs 145
          :::|||   :::::|||||||||         |||:::::::::::::::
     1207 TTCCACTTTACCAGCTGGCCAGACTTTGGGGTGCCTTTTACCCCGATCGG 1256

      145 pGlnLeuGluTyrLeuAlaAspArgValLysAsnSerAsnGlnAsnGlyA 162
          :   |||   :::|||      :::||||||:::::::|||::::::   |
     1257 CATGCTCAAGTTCCTC...AAGAAGGTGAAGGCCTGTAACCCTCAGTATG 1303

      162 laProGlyArgSerSerSerAspLysHisLeuProMetIleHisCysLeu 178
          ||   :::                          ::::::||||||
     1304 CAGGGGCCATC.......................GTGGTCCACTGCAGT 1329

      179 GlyGlyValGlyArgThrGlyThr 186
          :::||||||||||||||||||||
     1330 GCAGGTGTAGGGCGTACAGGTACC 1353

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-348-006B-6

seq_documentation_block:
; Sequence 6, Application US/08348006B
; Patent No. 5658756
;  GENERAL INFORMATION:
;    APPLICANT:  RODAN, GIDEON A.
;    APPLICANT:  SCHMIDT, AZRIEL
;    APPLICANT:  RUTLEDGE, SU JANE
;    TITLE OF INVENTION:  CDNA ENCODING A NOVEL HUMAN PROTEIN
;    TITLE OF INVENTION:  TYROSINE PHOSPHATASE
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  J. MARK HAND
;      STREET:  126 E. LINCOLN AVE., P.O. BOX 2000
;      CITY:  RAHWAY
;      STATE:  NEW JERSEY
;      COUNTRY:  USA
;      ZIP:  07065-0900
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/348,006B
;      FILING DATE:
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/122,032
;      FILING DATE:  14-SEP-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  HAND, J., MARK
;      REGISTRATION NUMBER:  36,545
;      REFERENCE/DOCKET NUMBER:  18992IA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  908-594-3905
;      TELEFAX:  908-594-4720
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6000 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-348-006B-6


alignment_scores:
            Quality:   84.50            Length:    255
              Ratio:  0.761              Gaps:      12
 Percent Similarity:  43.529   Percent Identity: 19.608

alignment_block:
 US-09-068-804-12 x US-08-348-006B-6   ..

 Align seg 1/1  to: US-08-348-006B-6  from: 1  to: 6000
```

```
     19 LysLeuSerAspGlyMetProValProValIleArgLeuThrPheAspGl 35
        :::|||   :::   ||||||                    ::::::::
   5039 CGCCTGGTGAACATCATGCCC.....................TATGAGAG 5067

     35 yLysProValAlaLeuAlaGlyIleArg..................... 44
        ::::   |||   |||      ||||||
   5068 CACACGGGTCTGTCTGCAACCCATCCGGGGTGTGGAGGGCTCTGACTACA 5117

     44 ..................................................... 44

   5118 TCAACGCCAGCTTCATTGATGGCTACAGGCAGCAGAAGGCCTACATCGCG 5167

     45 ThrGlnLysIleArgProAspArgLeuGluAlaHisMetLysMetLeuLe 61
        ||||||         ::::::      |||:::      :::||||||
   5168 ACACAGGGGCCGCTGGCGGAGACCACGGAAGACTTCTGGCGCATGCTGTG 5217

     61 uGluLysGluCysSerCysLeuValValLeuThrSerGluArgSerAspA 78
        |||:::::::::::   :::|||:::||||||      |||      :
   5218 GGAGAACAATTCGACGATCGTGGTGATGCTGACCAAGCTGCGGGAGATGG 5267

     78 laGlyLysThrIleThrThrTyrPhe.....ArgGlySerTyr 90
        ::      :::         |||                ::::  |||
   5268 GCCGGGAGAAGTGTCACCAGTACTGGCCGGCCGAGCGCTCTGCCCGCTAC 5317

     91 ThrPheGlyGluValHisThrAsnSerGlnLysValSerSerAlaSerGl 107
        :::      |||
   5318 CAGTACTTTGTGGTA..................................... 5332

    107 nGlyGluAlaIleAspGlnTyrAsnMetGlnLeuSerCys..... 120
        ::::::::::::::|||||||||:::
   5333 ....GATCCGATGGCAGAATACAACATGCCTCAGTATATCCTGCGAGAGT 5378

    121 .....................GlyGluLysArgTyrThrIleProVal 129
                             |||:::   |||   |||:::
   5379 TCAAGGTCACAGATGCCCGGGATGGCCAGTCCCGG...ACTGTCCGGCAG 5425

    130 LeuHisValLysAsnTrpProAspHisGlnProLeuProSerThrAspGl 146
        ::::::   ::::::||||||:::   |||   :::|||   ::::::::
   5426 TTCCAGTTCACAGACTGGCCGGAA...CAGGGTGTGCCAAAGTCGGGGGA 5472

    146 n......LeuGluTyrLeuAlaAspArgValLysAsnSerAsnGlnAsnG 161
        :       ::::::::::::::::::      |||:::   :::|||   |
   5473 GGGCTTCATCGACTTCATTGGCCAAGTGCATAAGACTAAGGAGCAGTTTG 5522

    161 lyAlaProGlyArgSerSerSerAspLysHisLeuProMetIleHisCys 177
        ||      |||       |||                 :::||||||
   5523 GCCAGGACGGCCCCATCTCT.....................GTCCACTGC 5551

    178 LeuGlyGlyValGlyArgThrGlyThrMetAlaAlaAlaLeuValLeuLy 194
        :::||||||||||||||||||:::      :::      ::::::
   5552 AGTGCCGGCGTGGGCAGGACGGGCGTCTTCATCACGCTTAGCATCGTG.. 5599

    194 sAspAsnProHisSerAsnLeuGluGlnValArgAlaAspPheArgIleH 211
                        ||||||:::::::|||   :::
   5600 ..................CTGGAGCGGATGCGGTATGAA.......... 5620

    211 isGlyThrIleGluCysTrpLysThrProAlaGlnPheValGlnLeuLys 227
          |||:::::::::                  |||      :::|||
   5621 ..GGCGTGGTGGACATC.................TTTCAGACGGTGAAG 5650

    228 AlaMetGlnAlaGln 232
        :::::::::|||
   5651 ATGCTACGAACCCAG 5665

seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-800-825A-6

seq_documentation_block:
; Sequence 6, Application US/08800825A
; Patent No. 5866397
;  GENERAL INFORMATION:
;    APPLICANT:  RODAN, GIDEON A.
;    APPLICANT:  SCHMIDT, AZRIEL
;    APPLICANT:  RUTLEDGE, SU JANE
;    TITLE OF INVENTION:  CDNA ENCODING A NOVEL HUMAN PROTEIN
;    TITLE OF INVENTION:  TYROSINE PHOSPHATASE
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  J. MARK HAND - MERCK & CO., INC.
;      STREET:  126 E. LINCOLN AVE., P.O. BOX 2000
;      CITY:  RAHWAY
;      STATE:  NEW JERSEY
;      COUNTRY:  USA
;      ZIP:  07065-0900
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/800,825A
;      FILING DATE:  14-FEB-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  HAND, J. MARK
;      REGISTRATION NUMBER:  36,545
;      REFERENCE/DOCKET NUMBER:  18992DA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  732-594-3905
;      TELEFAX:  732-594-4720
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6000 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-800-825A-6


alignment_scores:
            Quality:   84.50              Length:    255
              Ratio:  0.761                Gaps:     12
 Percent Similarity:  43.529   Percent Identity: 19.608

alignment_block:
 US-09-068-804-12 x US-08-800-825A-6   ..

 Align seg 1/1  to: US-08-800-825A-6  from: 1  to: 6000

      19 LysLeuSerAspGlyMetProValProValIleArgLeuThrPheAspGl 35
         ::::||   :::   ||||||                    ::::::::
    5039 CGCCTGGTGAACATCATGCCC.....................TATGAGAG 5067

      35 yLysProValAlaLeuAlaGlyIleArg..................... 44
         ::::   |||   |||      ||||||
    5068 CACACGGGTCTGTCTGCAACCCATCCGGGGTGTGGAGGGCTCTGACTACA 5117

      44 ..................................................... 44

    5118 TCAACGCCAGCTTCATTGATGGCTACAGGCAGCAGAAGGCCTACATCGCG 5167

      45 ThrGlnLysIleArgProAspArgLeuGluAlaHisMetLysMetLeuLe 61
         ||||||         ::::::      |||:::      :::||||||
    5168 ACACAGGGGCCGCTGGCGGAGACCACGGAAGACTTCTGGCGCATGCTGTG 5217

      61 uGluLysGluCysSerCysLeuValValLeuThrSerGluArgSerAspA 78
         |||:::::::::::   :::|||:::||||||      |||      :
    5218 GGAGAACAATTCGACGATCGTGGTGATGCTGACCAAGCTGCGGGAGATGG 5267

      78 laGlyLysThrIleThrThrTyrPhe.....ArgGlySerTyr 90
         ::      :::         |||                ::::  |||
    5268 GCCGGGAGAAGTGTCACCAGTACTGGCCGGCCGAGCGCTCTGCCCGCTAC 5317

      91 ThrPheGlyGluValHisThrAsnSerGlnLysValSerSerAlaSerGl 107
         :::      |||
```

```
5318 CAGTACTTTGTGGTA..................................... 5332

 107 nGlyGluAlaIleAspGlnTyrAsnMetGlnLeuSerCys.......... 120
        :::::::::::::::|||||||||:::
5333 ....GATCCGATGGCAGAATACAACATGCCTCAGTATATCCTGCGAGAGT 5378

 121 ........................GlyGluLysArgTyrThrIleProVal 129
                              |||:::   |||   |||:::
5379 TCAAGGTCACAGATGCCCGGGATGGCCAGTCCCGG...ACTGTCCGGCAG 5425

 130 LeuHisValLysAsnTrpProAspHisGlnProLeuProSerThrAspGl 146
     ::::::   ::::::||||||:::   |||   :::|||   ::::::::
5426 TTCCAGTTCACAGACTGGCCGGAA...CAGGGTGTGCCAAAGTCGGGGGA 5472

 146 n......LeuGluTyrLeuAlaAspArgValLysAsnSerAsnGlnAsnG 161
     :      ::::::::::::::::::      |||:::   :::|||   |
5473 GGGCTTCATCGACTTCATTGGCCAAGTGCATAAGACTAAGGAGCAGTTTG 5522

 161 lyAlaProGlyArgSerSerSerAspLysHisLeuProMetIleHisCys 177
     ||      |||      |||                    :::||||||
5523 GCCAGGACGGCCCCATCTCT.....................GTCCACTGC 5551

 178 LeuGlyGlyValGlyArgThrGlyThrMetAlaAlaAlaLeuValLeuLy 194
        :::||||||||||||||||||:::      :::      ::::::
5552 AGTGCCGGCGTGGGCAGGACGGGCGTCTTCATCACGCTTAGCATCGTG.. 5599

 194 sAspAsnProHisSerAsnLeuGluGlnValArgAlaAspPheArgIleH 211
                        ||||||::::::|||   :::
5600 ...................CTGGAGCGGATGCGGTATGAA.......... 5620

 211 isGlyThrIleGluCysTrpLysThrProAlaGlnPheValGlnLeuLys 227
       |||:::::::::                  |||      :::|||
5621 ..GGCGTGGTGGACATC.................TTTCAGACGGTGAAG 5650

 228 AlaMetGlnAlaGln 232
        ::::::::::|||
5651 ATGCTACGAACCCAG 5665
```

OM of: US-09-068-804-12 to: N_Geneseq_36:* out_format : pfs

Date: Oct 5, 1999 8:04 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=x1p
-Q=/cgn2_1/USPTO_spool/US09068804/runat_05101999_145135_10341/app_query.fasta.1
-DB=N_Geneseq_36 -QFMT=fastap -SUFFIX=rng -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=pam150
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=escore
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09068804
-NCPU=6 -ICPU=3 -WAIT -THREADS=1

Search information block:
Query: US-09-068-804-12
Query length: 239
Database: N_Geneseq_36:*
Database sequences: 311585
Database length: 125096042
Search time (sec): 156.120000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ..
N_Geneseq_36:T67033	-	819.00	1935.78	3.2e-100	579	! Salmonella tyrosine phosphatase
N_Geneseq_36:T37466	-	123.00	255.59	1.2e-06	2405	! Salmonella enterica IagA and I
N_Geneseq_36:V33577	+	92.00	172.54	0.0520	5455	! Osteotesticular protein tyrosi
N_Geneseq_36:V15004	+	89.00	164.76	0.1412	5769	! Receptor protein tyrosine phos
N_Geneseq_36:T35307	+	84.50	164.11	0.1535	2276	! Human SH-PTP2 gene (protein ty
N_Geneseq_36:V50490	+	79.00	160.75	0.2361	924	! Human cystatin-like protein enc
N_Geneseq_36:Q45353	+	84.00	160.70	0.2378	2790	! Human protein-tyrosine-phospha
N_Geneseq_36:Q13802	+	83.00	160.47	0.2449	2287	! Non-receptor linked protein ty
N_Geneseq_36:V71075	+	82.50	158.16	0.3290	2532	! Green fluorescent protein-prot
N_Geneseq_36:V71076	+	82.50	158.04	0.3345	2562	! Protein tyrosine phosphatase 1
N_Geneseq_36:V04868	+	84.50	156.10	0.4285	4752	! Nucleotide sequence of human g
N_Geneseq_36:V04887	+	84.50	156.00	0.4344	4799	! Nucleotide sequence of a human
N_Geneseq_36:V04886	+	84.50	155.96	0.4365	4815	! Nucleotide sequence of a human
N_Geneseq_36:T48479	+	84.50	155.31	0.4746	5113	! Human tyrosine dephosphorylase
N_Geneseq_36:Q21001	+	81.50	154.40	0.5330	2872	! Murine receptor-type protein t
N_Geneseq_36:T36881	+	81.50	154.40	0.5330	2872	! Murine receptor type protein t
N_Geneseq_36:X29853	+	81.50	154.40	0.5330	2872	! Mouse receptor-type phosphotyr
N_Geneseq_36:T36882	+	80.50	153.92	0.5669	2409	! Human receptor type protein ty
N_Geneseq_36:X29852	+	80.50	153.92	0.5669	2409	! Human receptor-type phosphotyr
N_Geneseq_36:Q86478	+	84.50	153.57	0.5931	6000	! Human PTP-OB. Protein tyrosine
N_Geneseq_36:T85389	+	84.50	153.57	0.5931	6000	! Human protein tyrosine phospha
N_Geneseq_36:X06095	+	84.50	153.57	0.5931	6000	! Human protein tyrosine phospha
N_Geneseq_36:V17098	+	83.00	150.77	0.8495	5581	! Human pancreatic carcinoma phc
N_Geneseq_36:T93334	+	72.00	150.47	0.8828	510	! EST locus ze82b09.r1 from cDNA
N_Geneseq_36:V82821	+	80.50	149.49	1.00	3623	! Rabbit osteoclast-specific phc
N_Geneseq_36:T35306	+	78.00	149.20	1.04	2145	! Human SH-PTP1 gene (protein ty
N_Geneseq_36:Q94311	+	81.00	148.19	1.18	4555	! Tyrosine phosphatase MPTP-delt
N_Geneseq_36:V34371	+	72.50	145.76	1.62	878	! PCR fragment of human protein t
N_Geneseq_36:V04867	+	80.50	144.60	1.87	5679	! Nucleotide sequence of rabbit
N_Geneseq_36:I58627	+	78.50	143.67	2.11	3984	! Protein tyrosine phosphatase c
N_Geneseq_36:X15661	+	76.00	141.57	2.76	2787	! Protein phosphatase gene codin
N_Geneseq_36:T35308	+	73.50	140.67	3.10	1747	! Rat M1PTP gene cDNA clone deri
N_Geneseq_36:T36035	+	78.50	138.51	4.09	6407	! rchd528 gene differentially ex
N_Geneseq_36:V81825	+	78.50	138.51	4.09	6407	! Human rchd528 encoding cDNA. D
N_Geneseq_36:X26246	+	78.50	138.51	4.09	6407	! cDNA sequence of rchd528 gene.
N_Geneseq_36:V24010	+	73.50	138.38	4.16	2157	! Human FK506 binding protein, F
N_Geneseq_36:V45137	+	72.00	137.22	4.83	1725	! Homo sapiens P-TEN tumour supp
N_Geneseq_36:T93354	+	72.00	136.92	5.02	1773	! Gene of IMAGE clone 264611. Nu
N_Geneseq_36:X07339	+	72.00	136.86	5.06	1784	! Dual specificity phosphatase P
N_Geneseq_36:Q73785	+	66.50	136.86	5.06	532	! Partial PCR fragment of PTP-S31
N_Geneseq_36:Q20381	+	68.00	135.75	5.83	819	! Sequence encoding the shortened
N_Geneseq_36:V81742	+	74.50	135.27	6.20	3580	! Human PTP04 encoding cDNA. New
N_Geneseq_36:Q70152	+	74.00	134.89	6.51	3319	! K39 gene of Leishmania chag
N_Geneseq_36:T42166	+	74.00	134.89	6.51	3319	! Leishmania chagasi K39 anti
N_Geneseq_36:T64586	+	68.50	134.79	6.60	999	! Modified human p65 protein c

seq_name: N_Geneseq_36:T67033

seq_documentation_block:
ID T67033 standard; DNA; 579 BP.
AC T67033;
DT 04-AUG-1997 (first entry)
DE Salmonella tyrosine phosphatase stpA gene.
KW StpA; tyrosine phosphatase; phosphorylation; signal transduction;
KW Salmonella secreted protein; Ssp; bacterial-mediated endocytosis;
KW ss.
OS Salmonella typhimurium.
FH Key Location/Qualifiers
FT cds 1. .579
FT /*tag= a
FT /product= StpA
FT /note= "the coding sequence does not translate
FT to the StpA amino acid sequence of Fig 25"
PN WO9718225-A1.
PD 22-MAY-1997.
PF 14-NOV-1996; U18504.
PR 14-NOV-1995; US-006733.
PA (GEHO) GEN HOSPITAL CORP.
PI Miller SI;
DR WPI; 97-289217/26.
PT New isolated Salmonella secreted proteins and related genes - used
PT to develop products for the detection, treatment or prevention of
PT Salmonella infections
PS Claim 15; Fig 24; 95pp; English.
CC Salmonella typhimurium stpA DNA (T67033) codes for Salmonella
CC tyrosine phosphatase (StpA) (W15273), an enzyme that catalyses the
CC release of phosphate groups from tyrosine residues in proteins.
CC StpA can be used to deactivate or activate eukaryotic or
CC prokaryotic signal transduction proteins that are regulated by the
CC phosphorylation and dephosphorylation of tyrosine residues, thereby
CC altering intracellular signal transduction. Thus, Stp can be used
CC as a research tool to study and evaluate phosphorylation-regulated
CC signal transduction pathways.
SQ Sequence 579 BP; 120 A; 158 C; 124 G; 177 T;

alignment_scores:
Quality: 819.00 Length: 191
Ratio: 4.356 Gaps: 0
Percent Similarity: 98.429 Percent Identity: 98.429

alignment_block:
US-09-068-804-12 x T67033/rev ..

Align seg 1/1 to reverse of: T67033 from: 1 to: 579

```
      49 ArgProAspArgLeuGluAlaHisMetLysMetLeuLeuGluLysGluCy 65
         |||  |||  ||||||||||||||||||||||||||||||||||||||||
     575 CGCCGGGATGCGCTGGAAGCGCATATGAAAATGCTTCTTGAAAAAGAATG 526

      65 sSerCysLeuValValLeuThrSerGluArgSerAspAlaGlyLysThrI 82
         |||||||||||||||||||||||  |||||||||||||||||||||||||
     525 CTCGTGCCTGGTGGTGTTAACGTCGGA.AGATCAGATGCAGGCAAAACAA 477

      82 leThrThrTyrPheArgGlySerTyrThrPheGlyGluValHisThrAsn 98
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     476 TTACCACCTATTTCAGAGGGAGCTACACCTTTGGCGAGGTGCATACCAAC 427

      99 SerGlnLysValSerSerAlaSerGlnGlyGluAlaIleAspGlnTyrAs 115
         |||||||||  ||||||||||||||||||||||||||||||||||||||||
     426 AGCCAAAAAGAGAGCTCAGCCAGTCAGGGAGAAGCGATAGACCAATACAA 377

     115 nMetGlnLeuSerCysGlyGluLysArgTyrThrIleProValLeuHisV 132
         |||||||||||||||||  |||||||||||||||||||||||||||||||
     376 TATGCAACTGTCCTGCGGGGA.AAGCGGTATACAATCCCGGTATTGCATG 328
```

```
    132 alLysAsnTrpProAspHisGlnProLeuProSerThrAspGlnLeuGlu 148
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    327 TGAAAAATTGGCCAGATCACCAGCCGTTACCGTCTACGGATCAGTTAGAA 278

    149 TyrLeuAlaAspArgValLysAsnSerAsnGlnAsnGlyAlaProGlyAr 165
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    277 TACCTGGCGGATAGGGTGAAAAATAGTAACCAAAATGGCGCGCCGGGGCG 228

    165 gSerSerSerAspLysHisLeuProMetIleHisCysLeuGlyGlyValG 182
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    227 TAGTTCATCAGATAAGCATTTACCGATGATTCATTGTCTGGGCGGAGTGG 178

    182 lyArgThrGlyThrMetAlaAlaAlaLeuValLeuLysAspAsnProHis 198
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    177 GAAGAACCGGAACGATGGCGGCGGCCCTTGTACTTAAGGATAATCCTCAT 128

    199 SerAsnLeuGluGlnValArgAlaAspPheArgIleHisGlyThrIleGl 215
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    127 AGTAATCTGGAGCAGGTACGTGCAGATTTCCGGATTCACGGAACAATAGA 78

    215 uCysTrpLysThrProAlaGlnPheValGlnLeuLysAlaMetGlnAlaG 232
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     77 ATGCTGGAAGACGCCTGCACAGTTTGTACAACTAAAAGCAATGCAAGCCC 28

    232 lnLeuLeuMetThrThrAlaSer 239
        |||||||||||||||||||||||
     27 AGTTGCTTATGACGACGGCAAGC 5
```

seq_name: N_Geneseq_36:T37466

seq_documentation_block:

```
ID   T37466 standard; DNA; 2405 BP.
AC   T37466;
DT   25-FEB-1997  (first entry)
DE   Salmonella enterica IagA and IagB operon sequences.
KW   IagA; IagB; Salmonella enterica; primer; probe; HeLa; ss.
OS   Salmonella enterica.
FH   Key             Location/Qualifiers
FT   cds             97. .1758
FT                   /*tag=  a
FT                   /product= IagA protein
FT   misc_signal     1769. .1772
FT                   /*tag=  b
FT                   /note= "putative ribosome binding site"
FT   cds             1776. .2258
FT                   /*tag=  c
FT                   /product= IagB protein
PN   EP-721989-A1.
PD   17-JUL-1996.
PF   15-JAN-1996; 400098.
PR   16-JAN-1995; FR-000410.
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PA   (INSP ) INST PASTEUR.
PI   Le GUERN FELLOUS M,  Popoff MY;
DR   WPI; 96-322837/33.
DR   P-PSDB; W03528, W09225.
PT   New nucleic acid of S.enterica ssp. enterica involved in cell
PT   invasion  - and derived oligo-nucleotide(s) useful as primers and
PT   probes for detecting Salmonella in food etc.
PS   Claim 1; Fig 1; 33pp; French.
CC   Oligonucleotides T37451-65 are derived from the newly isolated sequence
CC   of the IagA and IagB sequences from Salmonella enterica ssp. enterica
CC   serovar typhi.  The IagA and IagB sequences (T37466) encode proteins
CC   which are involved in the invasion of cultured HeLa cells by S. enterica.
SQ   Sequence   2405 BP;    735 A;    488 C;    519 G;    663 T;
```

```
alignment_scores:
            Quality:  123.00            Length:     30
              Ratio:  4.241               Gaps:      0
 Percent Similarity:  96.667   Percent Identity: 96.667
```

```
alignment_block:
 US-09-068-804-12 x T37466/rev   ..

 Align seg 1/1  to reverse of: T37466  from: 1  to: 2405

     210 IleHisGlyThrIleGluCysTrpLysThrProAlaGlnPheValGlnLe 226
         ||||||||||||||||||||||||||||||||  ||||||||||||||
    2403 ATTCACGGAACAATAGAATGCTGGAAGACGCCTCA.CAGTTTGTACAACT 2355

     226 uLysAlaMetGlnAlaGlnLeuLeuMetThrThrAlaSer 239
         |||||||||||||||||||||||||||||||||||||||
    2354 AAAAGCAATGCAAGCCCAGTTGCTTATGACGACGGCAAGC 2315
```

seq_name: N_Geneseq_36:V33577

seq_documentation_block:

```
ID   V33577 standard; DNA; 5455 BP.
AC   V33577;
DT   29-DEC-1998  (first entry)
DE   Osteotesticular protein tyrosine phosphatase cDNA.
KW   Osteotesticular protein tyrosine phosphatase; OST-PTP; bone; testis;
KW   osteoblast specific form; OST; OST-PTP receptor; asteoporosis;
KW   osteopetrosis; bone metabolic disorder; ss.
OS   Rattus sp.
FH   Key             Location/Qualifiers
FT   CDS             205. .5340
FT                   /*tag=  a
FT                   /product= "OST-PTP"
FT   polyA_signal    5449
FT                   /*tag=  b
PN   US5821084-A.
PD   13-OCT-1998.
PF   21-NOV-1994; 342930.
PR   21-NOV-1994; US-342930.
PA   (AMHP ) AMERICAN HOME PROD CORP.
PA   (UNMI ) UNIV MICHIGAN.
PI   Davis AR, Dixon JE, Mauro LJ, Olmsted EA;
DR   WPI; 98-567659/48.
DR   P-PSDB; W70506.
PT   DNA encoding osteoblast-testicular protein tyrosine kinase
PT   polypeptide - useful for producing recombinant polypeptide
PS   Claim 1; Columns 23-24; 31pp; English.
CC   The present sequence represents a rat osteotesticular protein tyrosine
CC   phosphatase (OST-PTP) cDNA which was isolated from the fetal rat
CC   calvaria and the rat osteosarcoma cell line, UMR106.  OST-PTP is a
CC   transmembrane protein expressed only in bone and testis.  The OST-PTP
CC   cDNA is useful for producing recombinant protein.  The invention also
CC   provides a truncated osteoblast specific form (OST) containing the
CC   OST-PTP receptor but lacking the catalytic domain.  The OST-PTP protein
CC   is claimed to be useful in assays used for screening abnormal bone
CC   growth patterns and metabolic bone diseases.  The OST-PTP protein is
CC   also claimed to be useful in the treatment of asteoporosis, osteopetrosis
CC   and other bone metabolic disorders.
SQ   Sequence   5455 BP;    1169 A;    1607 C;    1525 G;    1154 T;
```

```
alignment_scores:
            Quality:   92.00            Length:    214
              Ratio:  0.800               Gaps:      9
 Percent Similarity:  53.738   Percent Identity: 20.093
```

```
alignment_block:
 US-09-068-804-12 x V33577   ..

 Align seg 1/1  to: V33577  from: 1  to: 5455

      17 GlnValLysLeuSerAspGlyMetProValProValIleArg..... 30
         |||   :::|||   :::      |||      :::|||:::
    3694 CAGCCCCGACTAGAGGCTGAGCATCCGGACAACATCATCAAGAACCGGTA 3743

      31 .........LeuThrPheAspGlyLysProValAlaLeuAlaGlyIle. 43
                  |||:::::::|||         |||   |||:::   :::
    3744 CCCACACGTGCTGCCCTATGACCACTCCAGGGTCAGGCTGACCCAGCTAC 3793
```

```
      43 ................................................ 43

    3794 CAGGAGAGCCTCATTCTGACTACATCAATGCCAACTTCATCCCAGGCTAT 3843

      44 ...ArgThrGlnLysIleArgProAspArg.................Le 53
           :::|||||||   |||   :::   :::                  ||
    3844 AGCCACACACAGGAGATCATTGCCACCCAGGGGCCTCTCAAAAAGACGCT 3893

      53 uGluAlaHisMetLysMetLeuLeuGluLysGluCysSerCysLeuValV 70
         ||||:::      :::::::::   |||::::::        :::::::
    3894 AGAGGACTTCTGGCGGTTGGTATGGGAGCAGCAAGTCCACGTGATCATCA 3943

      70 alLeuThrSerGluArgSerAspAlaGlyLysThrIleThrThrTyrPhe 86
         ::|||||||           :::::::|||:::::::::      ::::::
    3944 TGCTGACTGTGGGCATG...GAGAACGGGCGGGTACTGTGTGAGCACTAC 3990

      87 Arg......GlySerTyrThrPheGlyGluValHisThrAsnSerGlnLy 101
         :::      :::|||   :::      :::   |||::::::   :::
    3991 TGGCCAGCCAACTCCACGCCTGTTACTCACGGTCACATCACCATCCACCT 4040

     101 sValSerSerAlaSerGlnGlyGlu...AlaIleAspGlnTyrAsnMetG 117
          ::::::   :::::::::::::|||   :::      ::::::::::::|
    4041 CCTGGCAGAGGAGCCTGAGGATGAGTGGACCAGGAGGGAATTCCAGCTGC 4090

     117 lnLeuSerCysGlyGluLysArgTyrThrIleProValLeuHisValLys 133
         ||   :::      :::|||:::      :::      |||:::   :::
    4091 AGCACGGTACCGAGCAAAAACAGAGGCGAGTGAAGCAGCTGCAGTTCACT 4140

     134 AsnTrpProAspHisGln...ProLeuProSerThrAspGlnLeuGluTy 149
         :::|||||||||||||      |||   ::::::::::::   |||::::::
    4141 ACCTGGCCAGACCACAGTGTCCCGGAGGCTCCCAGCTCTCTGCTCGCTTT 4190

     149 rLeuAla.........AspArgValLysAsnSerAsnGlnAsnGlyAlaP 163
         :::::::         ::::::|||:::::::::::   :::|||:::
    4191 TGTAGAACTGGTACAGGAGCAGGTGCAGGCCACTCAGGGCAAGGGACCC. 4239

     163 roGlyArgSerSerSerAspLysHisLeuProMetIleHisCysLeuGly 179
                                        ::::::||||||   :::
    4240 ............................ATCCTGGTGCATTGCAGTGCT 4260

     180 GlyValGlyArgThrGlyThrMetAlaAlaAlaLeuValLeu 193
         ||||||||||||||||||||   :::|||   |||   |||
    4261 GGCGTGGGGAGGACAGGCACCTTTGTGGCTCTCTTGCGGCTA 4302

seq_name: N_Geneseq_36:V15004

seq_documentation_block:
ID   V15004 standard; DNA; 5769 BP.
AC   V15004;
DT   26-MAY-1998  (first entry)
DE   Receptor protein tyrosine phosphatase lambda polypeptide coding sequence.
KW   Receptor protein tyrosine phosphatase lambda; mouse; PTP lambda; therapy;
KW   cognate ligand; paralytic disease; metastatic spread prevention;
KW   epithelial development; neuronal structure development;
KW   neuronal pathfinding; ss.
OS   Mus sp.
FH   Key             Location/Qualifiers
FT   CDS             379. .4689
FT                   /*tag= a
FT                   transl_except= (pos:850. .852, aa:Xaa)
FT                   transl_except= (pos:1395. .1397, aa:Xaa)
FT                   transl_except= (pos:4000. .4002, aa:Xaa)
FT                   /note= "Xaa = unknown"
PN   WO9744458-A1.
PD   27-NOV-1997.
PF   22-MAY-1997; U09056.
PR   24-MAY-1996; US-652971.
PA   (GETH ) GENENTECH INC.
PI   Cheng J, Lasky LA;
DR   WPI; 98-018512/02.
DR   P-PSDB; W41361.
PT   Receptor protein tyrosine phosphatase lambda polypeptide -
PT   potentially useful in treating paralytic diseases and preventing
PT   metastases
PS   Example 1; Fig 1; 118pp; English.
CC   This sequence encodes the mouse receptor protein tyrosine phosphatase
CC   (PTP) lambda polypeptide of the invention. PTP lambda is: (i) mainly
CC   expressed in adult mammalian brain, lung and kidney; (ii) not expressed
CC   in adult liver; and (iii) can dephosphorylate phosphorylated tyrosine
CC   (pY) residues. The DNA sequence is used to express recombinant PTP
CC   lambda, and also as a source of probes for screening DNA libraries.
CC   PTP lambda is used to isolate the receptor's cognate ligand (potentially
CC   useful for treating paralytic diseases), and for preventing metastatic
CC   spread. PTP lambda is also used as a molecular marker for particular
CC   tissues, as a reagent in an assay for identifying PTP (ant)agonists, and
CC   as a molecular weight marker in protein gels. The Ab is used in
CC   diagnosis, tissue-typing and purification of PTP lambda, also it may
CC   crosslink the receptor, resulting in its upregulation. PTP lambda is a
CC   mediator of cell adhesion and may be involved in development of
CC   epithelial and neuronal structures (neuronal pathfinding).
SQ   Sequence   5769 BP;    1218 A;    1718 C;    1632 G;    1197 T;

alignment_scores:
            Quality:   89.00              Length:    151
              Ratio:  1.011                Gaps:      7
 Percent Similarity:  58.278   Percent Identity: 23.179

alignment_block:
 US-09-068-804-12 x V15004   ..

 Align seg 1/1  to: V15004  from: 1  to: 5769

      43 IleArgThrGlnLysIleArgProAspArgLeuGluAlaHisMetLysMe 59
         |||   ||||||      :::|||:::   :::   :::      :::||
    3223 ATAGCCACTCAAGGGCCAAAGCCTGAGATGATCTACGATTTCTGGCGCAT 3272

      59 tLeuLeuGluLysGluCysSerCysLeuValValLeuThrSerGluArgS 76
         |:::   :::   :::|||:::::   :::|||:::::::::
    3273 GGTGTGGCAGGAACAGTGTGCGAGC...ATCGTCATGATCACCAAGCTGG 3319

      76 erAspAlaGlyLysThrIleThrThrTyrPheArg......GlySerTyr 90
         ::::::|||:::::      :::   ::::::      :::|||
    3320 TAGAGGTGGGCAGGGTGAAGTGTTCTCGCTACTGGCCTGAGGACTCAGAC 3369

      91 ThrPheGlyGluValHisThrAsnSerGlnLysValSerSerAlaSerGl 107
         :::|||::::::   ::::::      |||:::
    3370 ATGTATGGGGACATCAAGATCACGCTGGTAAAGACA.............. 3405

     107 nGlyGluAlaIleAspGlnTyrAsnMetGlnLeuSerCysGlyGluLys. 123
         |||:::::::::::::|||   ::::::          |||:::
    3406 ....GAGACACTGGCTGAGTATGTGGTGCGCACCTTTGCCCTGGAGCGGA 3451

     124 .............ArgTyrThrIleProValLeuHisValLysAsnTrp 135
                      |||:::   :::      :::|||   ::::::|||
    3452 GAGGTTACTCAGCCCGGCATGAGGTCCGCCAGTTCCATTTCACAGCGTGG 3501

     136 ProAspHis...GlnProLeuProSerThrAspGlnLeuGluTyrLeuAl 151
         |||:::|||      |||      :::|||:::   |||:::::::::
    3502 CCAGAGCATGGTGTCCCCTACCACGCCACGGGGCTGCTGGCCTTCATC.. 3549

     151 aAspArgValLysAsnSerAsnGlnAsnGlyAlaProGlyArgSerSerS 168
         |||||||||:::|||::::::   :::|||
    3550 .CGGCGTGTGAAGGCTTCCACTCCACCTGATGCCGGGCCCATT....... 3591

     168 erAspLysHisLeuProMetIleHisCysLeuGlyGlyValGlyArgThr 184
                        :::||||||||||   :::|||:::||||||||||
    3592 ................GTCATTCACTGCAGTGCAGGAACTGGCCGCACA 3624

     185 Gly 185
         |||
    3625 GGC 3627

seq_name: N_Geneseq_36:T35307
```

```
seq_documentation_block:
ID   T35307 standard; cDNA; 2276 BP.
AC   T35307;
DT   13-NOV-1996  (first entry)
DE   Human SH-PTP2 gene (protein tyrosine phosphatase-2, with SH2 domains).
KW   PTP; protein tyrosine phosphatase; SH2; Src homology region 2;
KW   chromosome 12p; abnormality; mutation; detection; probe; neoplasia;
KW   cancer; leukaemia; diagnosis; megakaryocyte regulation; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             114. .1895
FT                   /*tag=  a
PN   US5536636-A.
PD   16-JUL-1996.
PF   26-JUN-1991; 721112.
PR   26-JUN-1991; US-721112.
PR   31-JAN-1992; US-829141.
PR   01-DEC-1992; US-983926.
PR   28-FEB-1994; US-202389.
PA   (BETH-) BETH ISRAEL HOSPITAL ASSOC.
PA   (MASI ) MASSACHUSETTS INST TECHNOLOGY.
PI   Freeman RM,  Neel BG,  Plutzky J,  Rosenberg RD;
DR   WPI; 96-341506/34.
DR   P-PSDB; R99313.
PT   Detecting 12p chromosomal abnormality associated with neoplastic
PT   disease - using SH-PTP1 protein tyrosine phosphatase gene specific
PT   probe
PS   Example 3; Column 53-58; 63pp; English.
CC   T35307 is a human cDNA sequence encoding SH-PTP2 (protein tyrosine
CC   phosphatase-2, with two SH2 domains). A complementary fragment of this
CC   sequence can be used as a probe to detect a chromosome 12p13
CC   abnormality associated with neoplastic disease, in partic. acute
CC   lymphoblastic leukaemia. The probe hybridises to a part of the region
CC   coding for the two tandem SH2 domains (see R99313). If the probe will
CC   not hybridise DNA of chromosome 12p13 from a patient sample it is
CC   indicative of an abnormality, normally associated with neoplasia.
CC   Fragments of sequences encoding human SH-PTP1 (see T35306) may also
CC   be used to diagnose a condition or susceptiblity to a condition
CC   associated with chromosome 12p abnormalities. Alternatively the
CC   wild-type SH-PTP1 or SH-PTP2 gene or protein may be used for comparison
CC   to sequenced PTP genes taken from a patient, where differences indicate
CC   an abnormality. The activity of SH-PTP1 may also be purposely altered
CC   by mutation to effect a change in megakaryocyte function and hence
CC   platelet production.
SQ   Sequence   2276 BP;    722 A;    458 C;    565 G;    531 T;


alignment_scores:
            Quality:   84.50              Length:    242
              Ratio:  0.710                Gaps:      9
 Percent Similarity:  49.174   Percent Identity: 20.248

alignment_block:
 US-09-068-804-12 x T35307   ..

 Align seg 1/1  to: T35307  from: 1  to: 2276

      14 GlnGlnThrGlnValLysLeuSerAspGlyMetPro..............  25
         ::::::|||:::|||   |||   ||||||   |||
     969 GATCATACCAGGGTTGTCCTACACGATGGTGATCCCAATGAGCCTGTTTC 1018

      26 .............ValProValIleArgLeuThrPheAspGlyLysProV  38
                      :::   :::|||         |||:::   |||
    1019 AGATTACATCAATGCAAATATCATCATGCCTGAATTTGAAACCAAGTGCA 1068

      38 alAlaLeuAlaGly............IleArgThrGlnLysIleArgPro  50
         :::                      |||   ||||||      :::
    1069 ACAATTCAAAGCCCAAAAAGAGTTACATTGCCACACAAGGCTGCCTGCAA 1118

      51 AspArgLeuGluAlaHisMetLysMetLeuLeuGluLysGluCysSerCy  67
         :::   :::::::::      :::|||:::::::::   ::::::
    1119 AACACGGTGAATGACTTTTGGCGGATGGTGTTCCAAGAAAACTCCCGAGT 1168

      67 sLeuValValLeuThrSerGluArgSerAspAlaGlyLysThrIleThrT  84
         :::|||:::   |||   |||         :::      :::   ::::
    1169 GATTGTCATGACAACGAAAGAAGTGGAGAGAGGAAAGAGTAAATGTGTCA 1218

      84 hrTyrPheArgGlySerTyrThr.........PheGlyGluValHisThr  97
         ::|||      :::   |||:::         :::|||   :::::::::
    1219 AATACTGGCCTGATGAGTATGCTCTAAAAGAATATGGCGTCATGCGTGTT 1268

      98 AsnSerGlnLysValSerSerAlaSerGlnGlyGluAlaIleAspGlnTy 114
         :::   |||   |||:::|||                       :::::||
    1269 AGGAACGTCAAAGAAAGCGCCGCT.................CATGACTA 1300

     114 rAsnMetGln.................LeuSerCysGlyGluLysArgT 125
         |:::::::::                 ::::::   |||::::::
    1301 TACGCTAAGAGAACTTAAACTTTCAAAGGTTGGACAAGGGAATACGGAGA 1350

     125 yrThrIleProValLeuHisValLysAsnTrpProAspHis...GlnPro 140
           |||:::         |||   ::::::|||||||||||||      |||
    1351 GAACGGTCTGGCAATACCACTTTCGGACCTGGCCGGACCACGGCGTGCCC 1400

     141 LeuProSerThrAspGlnLeuGluTyrLeuAlaAsp......ArgValLy 155
               :::   :::   |||:::::|||:::::      :::   ::
    1401 AGCGACCCTGGGGGCGTGCTGGACTTCCTGGAGGAGGTGCACCATAAGCA 1450

     155 sAsnSerAsnGlnAsnGlyAlaProGlyArgSerSerSerAspLysHisL 172
         ::::|||      ::::::::::|||
    1451 GGAGAGCATCATGGATGCAGGGCCG..........................  1475

     172 euProMetIleHisCysLeuGlyGlyValGlyArgThrGlyThrMetAla 188
              ::::::||||||   :::|||:::||||||||||||||||
    1476 ..GTCGTGGTGCACTGCAGTGCTGGAATTGGCCGGACAGGGACGTTCATT 1523

     189 AlaAlaLeuValLeuLysAsp............................  195
         :::      :::|||   |||
    1524 GTGATTGATATTCTTATTGACATCATCAGAGAGAAAGGTGTTGACTGCGA 1573

     196 ................................AsnProHisSerAsnL 201
                                        :::::::::|||::::
    1574 TATTGACGTTCCCAAAACCATCCAGATGGTGCGGTCTCAGAGGTCAGGGA 1623

     201 euGluGlnValArgAlaAspPheArg 209
         ::   |||:::   |||::::::|||
    1624 TGGTCCAGACAGAAGCACAGTACCGA 1649

seq_name: N_Geneseq_36:V50490

seq_documentation_block:
ID   V50490 standard; cDNA; 924 BP.
AC   V50490;
DT   26-OCT-1998  (first entry)
DE   Human cystatin-like protein encoding nucleotide sequence.
KW   Human; cystatin-like protein; CSTIN; diagnosis; cancer; infection;
KW   neurodegeneration; immunological disorder; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             200. .736
FT                   /*tag=  a
FT                   /product= "cystatin-like protein"
PN   WO9833921-A1.
PD   06-AUG-1998.
PF   30-JAN-1998; U01776.
PR   30-JAN-1997; US-791522.
PA   (INCY-) INCYTE PHARM INC.
PI   Bandman O, Goli SK;
DR   WPI; 98-437474/37.
DR   P-PSDB; W69734.
PT   New human cystatin-like protein - used for treatment, prevention and
PT   diagnosis of cancer, neurodegeneration, infections and immunological
PT   disorders
PS   Claim 5; Page 44; 57pp; English.
CC   The present sequence encodes human cystatin-like protein (CSTIN). Host
CC   cells comprising vectors containing nucleotide sequences encoding CSTIN
CC   are used to produce recombinant CSTIN. CSTIN and its agonists are used
```

```
CC   to treat or prevent diseases associated with expression of CSTIN, e.g.
CC   cancer (lung, brain, colon, bone, breast or pancreas), neurodegeneration
CC   (e.g. Huntington's disease or muscular dystrophy) and infections
CC   (bacterial, viral or parasitic, e.g. human immune deficiency virus,
CC   rota or corona viruses, Streptococcus or Trypanosoma cruzi), also
CC   immunological conditions such as asthma, rheumatoid arthritis,
CC   spondylarthritis, osteoarthritis or glomerulonephritis). CSTIN can also
CC   be used to screen for specific inhibitors, e.g. antisense, triplex-
CC   forming, ribozyme or untranslatable nonsense nucleic acid, potentially
CC   useful therapeutically. Fragments of the nucleotide sequence encoding
CC   CSTIN are used as hybridisation probes or amplification primers for
CC   detecting CSTIN encoding nucleotide sequences, for diagnosing or
CC   monitoring disease, also for chromosome mapping. Antibodies (Ab)
CC   specific for CSTIN can be used in diagnostic immunoassays,
CC   therapeutically, for affinity purification of CSTIN from natural sources
CC   and in competitive screening assays.
SQ   Sequence   924 BP;    224 A;    303 C;    220 G;    175 T;


alignment_scores:
           Quality:   79.00              Length:     95
             Ratio:  1.491                Gaps:      5
 Percent Similarity:  55.789   Percent Identity: 27.368

alignment_block:
 US-09-068-804-12 x V50490   ..

 Align seg 1/1  to: V50490  from: 1  to: 924

     128 ProValLeuHisValLysAsnTrpProAspHisGlnProLeuProSerTh 144
          ||||||||||      |||:::|||   :::|||:::||||||||||:::::
      87 CCCGTGCTGCCTGAGAAGGATTGGCACGGGCACAGACCACTGCCCCCACC 136

     144 rAspGlnLeuGluTyrLeuAlaAspArgValLysAsnSerAsnGlnAsnG 161
          ::::   |||   :::|||:::   ::::::::::::::::::|||
     137 TGCC...CTGCGCCATCTACCCAAGAAGGCTCGGCACGGGCACCAACCAC 183

     161 lyAlaProGlyArgSerSerSerAspLysHisLeuProMetIleHisCys 177
           :::|||      :::::::::   ::::::            ||||||
     184 TGCCTCCAACTGCCCCATGCTGCCTGAGAAGG.....CACTGC 221

     178 LeuGlyGlyValGlyArgThrGlyThrMetAla...AlaAlaLeuValLe 193
                      |||:::|||   :::   :::||||||   ||
     222 .................ACGGCCACCCCCAACTGCCCCGCACTGTCCCT 253

     193 uLysAspAsnProHisSerAsnLeuGluGlnValArgAlaAspPheArgI 210
          |   ::::::|||         ||||||
     254 ACCCGGGCAGCCATGCGAGCGGCTGGAACTC..................... 284

     210 leHisGlyThrIleGluCysTrpLysThrProAla 221
                     ||||||   ::::::|||
     285 .................TGCTGGCCTTCTGCTGCC 302

seq_name: N_Geneseq_36:Q45353

seq_documentation_block:
ID   Q45353 standard; DNA; 2790 BP.
AC   Q45353;
DT   09-OCT-1994  (first entry)
DE   Human protein-tyrosine-phosphatase-1D DNA.
KW   Protein-tyrosine-phosphatase; enzyme; disease diagnosis.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             130. .1911
FT                   /*tag= a
PN   WO9408017-A.
PD   14-APR-1994.
PF   06-OCT-1993; E02728.
PR   06-OCT-1992; US-956315.
PR   16-FEB-1993; US-018129.
PA   (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PI   Ullrich A,  Vogel W;
DR   WPI; 94-135583/16.
DR   P-PSDB; R52991.
PT   New protein tyrosine phosphatase (PTP) protein, PTP-ID - are
PT   useful for diagnosis and treatment of diseases associated with
PT   abnormal PTP-ID levels
PS   Claim 8; Page 64; 99pp; English.
CC   This DNA may be expressed in a prokaryotic or eukaryotic host for
CC   the production of a PTP-1D protein.
SQ   Sequence   2790 BP;    861 A;    548 C;    669 G;    712 T;


alignment_scores:
           Quality:   84.00              Length:    243
             Ratio:  0.694                Gaps:      9
 Percent Similarity:  49.794   Percent Identity: 19.753

alignment_block:
 US-09-068-804-12 x Q45353   ..

 Align seg 1/1  to: Q45353  from: 1  to: 2790

       4 ThrIleCysGlyGlnGlnHisSerIleAsnGlnGlnThrGlnValLysLe 20
         |||||||||:::::::::::::::::::
    1006 ACGATCTGTGATCCCAATGAGCCTGTT..................... 1032

      20 uSerAspGlyMetProValProValIleArgLeuThrPheAspGlyLysP 37
          ||||||   :::   :::   :::|||         |||:::   |||
    1033 .TCAGATTACATCAATGCAAATATCATCATGCCTGAATTTGAAACCAAGT 1081

      37 roValAlaLeuAlaGly...........IleArgThrGlnLysIleArg 49
             :::                    |||   ||||||
    1082 GCAACAATTCAAAGCCCAAAAAGAGTTACATTGCCACACAAGGCTGCCTG 1131

      50 ProAspArgLeuGluAlaHisMetLysMetLeuLeuGluLysGluCysSe 66
          ::::::   :::::::::      :::|||:::::::::   ::::::
    1132 CAAAACACGGTGAATGACTTTTGGCGGATGGTGTTCCAAGAAAACTCCCG 1181

      66 rCysLeuValValLeuThrSerGluArgSerAspAlaGlyLysThrIleT 83
           :::|||:::   |||   |||       :::      :::   :
    1182 AGTGATTGTCATGACAACGAAAGAAGTGGAGAGAGGAAAGAGTAAATGTG 1231

      83 hrThrTyrPheArgGlySerTyrThr.........PheGlyGluValHis 96
          ::   |||      :::   |||:::         :::|||   ::::::
    1232 CTCAATACTGGCCTGATGAGTATGCTCTAAAAGAATATGGCGTCATGCGT 1281

      97 ThrAsnSerGlnLysValSerSerAlaSerGlnGlyGluAlaIleAspGl 113
          :::   :::   |||   |||:::|||                  :::::
    1282 GTTAGGAACGTCAAAGAAAGCGCCGCT..................CATGA 1313

     113 nTyrAsnMetGln..................LeuSerCysGlyGluLysA 124
          :|||::::::::::              ::::::   |||::::::
    1314 CTATACGCTAAGAGAACTTAAACTTTCAAAGGTTGGACAAGGGAATACGG 1363

     124 rgTyrThrIleProValLeuHisValLysAsnTrpProAspHis...Gln 139
             |||:::          |||   :::::||||||||||||
    1364 AGAGAACGGTCTGGCAATACCACTTTCGGACCTGGCCGGACCACGGCGTG 1413

     140 ProLeuProSerThrAspGlnLeuGluTyrLeuAlaAsp......ArgVa 154
          |||      :::   :::   |||::::::|||::::::      :::
    1414 CCCAGCGACCCTGGGGGCGTGCTGGACTTCCTGGAGGAGGTGCACCATAA 1463

     154 lLysAsnSerAsnGlnAsnGlyAlaProGlyArgSerSerSerAspLysH 171
          ::::::|||      ::::::::::|||
    1464 GCAGGAGAGCATCATGGATGCAGGGCCG..................... 1491

     171 isLeuProMetIleHisCysLeuGlyGlyValGlyArgThrGlyThrMet 187
             ::::::||||||   :::|||:::|||||||||||||||||
    1492 .....GTCGTGGTGCACTGCAGTGCTGGAATTGGCCGGACAGGGACGTTC 1536

     188 AlaAlaAlaLeuValLeuLysAsp........................... 195
           :::      :::|||  |||
    1537 ATTGTGATTGATATTCTTATTGACATCATCAGAGAGAAACGTGTTGACTG 1586

     196 .....................................AsnProHisSerA 200
```

```
                                                  :::::::::|||:
   1587 CGATATTGACGTTCCCAAAACCATCCAGATGGTGCGGTCTCAGAGGTCAG 1636

     200 snLeuGluGlnValArgAlaAspPheArg 209
         :::::   |||:::   |||::::::|||
   1637 GGATGGTCCAGACAGAAGCACAGTACCGA 1665
```

seq_name: N_Geneseq_36:Q13802

seq_documentation_block:

```
ID   Q13802 standard; cDNA; 2287 BP.
AC   Q13802;
DT   09-DEC-1991  (first entry)
DE   Non-receptor linked protein tyrosine phosphatase.
KW   PTPase; malignancy; cancer; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             61. .1308
FT                   /*tag=  a
FT                   /product= PTPase
FT   polya_signal    1512. .1526
FT                   /*tag=  b
FT   polya_signal    1677. .1682
FT                   /*tag=  c
PN   WO9113989-A.
PD   19-SEP-1991.
PF   14-MAR-1991; U01748.
PR   14-MAR-1990; US-494036.
PA   (WASH-) WASHINGTON RES FOUN.
PI   Fischer EH, Krebs EG, Tonks NK, Cool DE;
DR   WPI; 91-295643/40.
DR   P-PSDB; R14114.
PT   New DNA encoding non-receptor-linked protein tyrosine phosphatase
PT   - used in treating malignancies associated with protein-linked
PT   tyrosine phosphatase.
PS   Claim 1; Fig 1; 38pp; English.
CC   A cDNA library prepd. from human peripheral T cell poly(A)+ mRNA was
CC   screened using probes designed from the sequence of human placenta
CC   PTPase 1B.   The sequence can be used to express recombinant PTPase
CC   or derivs. useful in cancer therapy where a protein kinase is
CC   involved.
CC   See also Q13817-Q13820.
SQ   Sequence   2287 BP;    753 A;    411 C;    464 G;    659 T;
```

alignment_scores:

```
            Quality:   83.00            Length:    123
              Ratio:  1.064              Gaps:      9
 Percent Similarity:  63.415   Percent Identity: 30.894
```

alignment_block:

US-09-068-804-12 x Q13802 ..

Align seg 1/1 to: Q13802 from: 1 to: 2287

```
     114 TyrAsnMetGlnLeu.................SerCysGlyGluLysAr 124
         |||::::::::::|||                 ::::::||||||:::||
     520 TATACAGTACATCTACTACAATTAGAAAATATCAATAGTGGTGAAACCAG 569

     124 gTyrThrIleProValLeuHisValLysAsnTrpProAspHis...GlnP 140
         |   |||||||:::   :::|||   ::::::|||||||||        |
     570 A...ACAATATCTCACTTTCATTATACTACCTGGCCAGATTTTGGAGTCC 616

     140 roLeuProSerThrAspGlnLeuGluTyrLeuAlaAspArgValLysAsn 156
         ||   ::::::::::::   |||:::::::|||      ::::||:::::
     617 CTGAATCACCAGCTTCATTTCTCAATTTCTTGTTT...AAAGTGAGAGAA 663

     157 SerAsnGlnAsnGlyAlaProGlyArgSerSerSerAspLysHisLeuPr 173
         |||                 |||      ::::::|||         ::
     664 TCT.................GGCTCCTTGAACCCTGACCATGGGCCTGC 695

     173 oMetIleHisCysLeuGlyGlyValGlyArgThrGlyThr.........M 187
         ::::|||||||||   :::|||:::|||||||:::||||||         :
```

```
     696 GGTGATCCACTGTAGTGCAGGCATTGGGCGCTCTGGCACCTTCTCTCTGG 745

     187 etAlaAlaAlaLeuValLeuLysAspAsnProHisSer...AsnLeuGlu 202
         ::::::::   |||||||||:::::::::   ::::::   |||:::
     746 TAGACACTTGTCTTGTTTTGATGGAAAAAGGAGATGATATTAACATAAAA 795

     203 GlnValArgAlaAspPheArgIleHis......GlyThrIleGluCysTr 217
         ||||||      :::   |||   :::      |||   |||:::
     796 CAAGTGTTACTGAACATGAGAAAATACCGAATGGGTCTTATTCAG..... 840

     217 pLysThrProAlaGlnPhe 223
             ||||||:::|||:::
     841 ....ACCCCAGATCAACTG 855
```

seq_name: N_Geneseq_36:V71075

seq_documentation_block:

```
ID   V71075 standard; cDNA; 2532 BP.
AC   V71075;
DT   08-FEB-1999  (first entry)
DE   Green fluorescent protein-protein tyrosine phosphatase 1C fusion product.
KW   Human; protein tyrosine phosphatase 1C gene; fusion protein;
KW   green fluorescent protein; GFP; intracellular signalling; chimera; ss.
OS   Chimeric - Aequorea victoria.
OS   Chimeric - Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             1. .2532
FT                   /*tag=  a
PN   WO9845704-A2.
PD   15-OCT-1998.
PF   07-APR-1998; DK0145.
PR   07-APR-1997; DK-000392.
PA   (NOVO ) NOVO-NORDISK AS.
PI   Kasper A, Petersen Bjorn S, Scudder K, Thastrup O,
PI   Tullin S;
DR   WPI; 98-594491/50.
DR   P-PSDB; W85030.
PT   Determining effect on signalling pathways in live cells from
PT   redistribution of luminophores - specifically fusions of green
PT   fluorescent protein with a signalling component, and new apparatus,
PT   particularly for identifying toxins and potential therapeutic agents
PS   Claim 63; Pages 213-217; 326pp; English.
CC   The present sequence encodes a green fluorescent protein (GFP)-human
CC   protein tyrosine phosphatase 1C fusion protein. The fusion protein is
CC   used in an assay to exemplify the invention. The specification describes
CC   how quantitative information about the influence of a molecule on a
CC   cellular response is obtained by recording the variation, caused by
CC   the molecule, on mechanically intact living cells, in the spatially
CC   distributed light emitted from a luminophore present in the cells. The
CC   variation in light emission is processed to provide information that
CC   correlates spatial distribution to the degree of the molecule. The method
CC   is used to identify agents that (in)directly affect intracellular
CC   signalling, especially to screen for potential therapeutic agents or
CC   toxins, and to identify new drug targets.
SQ   Sequence   2532 BP;    613 A;    742 C;    758 G;    419 T;
```

alignment_scores:

```
            Quality:   82.50            Length:    167
              Ratio:  1.058              Gaps:      10
 Percent Similarity:  46.707   Percent Identity: 22.156
```

alignment_block:

US-09-068-804-12 x V71075 ..

Align seg 1/1 to: V71075 from: 1 to: 2532

```
     119 SerCysGlyGlu.........LysArgTyrThrIleProValLeuHisVa 132
         :::||||||||||         :::   |||::::::   :::|||:::||
    1885 AACTGCGGGGAGCATGACACAACCGAATACAAACTCCGTACCTTACAGGT 1934

     132 lLys..................................................... 133
         |
```

```
 1935 CTCCCCGCTGGACAATGGAGACCTGATTCGGGAGATCTGGCATTACCAGT 1984

  134 .....AsnTrpProAspHis...GlnProLeuProSerThrAspGlnLeu 147
           :::|||||||||||||      |||      :::   :::   |||
 1985 ACCTGAGCTGGCCCGACCATGGGGTCCCCAGTGAGCCTGGGGGTGTCCTC 2034

  148 GluTyrLeuAlaAspArgValLysAsnSerAsnGlnAsnGlyAlaProGl 164
        :::|||   |||:::::::::::   :::::::::
 2035 AGCTTCCTG...GACCAGATCAACCAGCGGCAGGAAAGT.......... 2070

  164 yArgSerSerSerAspLysHisLeuPro.............MetIleH 176
                          ||||||               ::::::|
 2071 ....................CTGCCTCACGCAGGGCCCATCATCGTGC 2098

  176 isCysLeuGlyGlyValGlyArgThrGlyThrMetAlaAlaAlaLeuVal 192
      |||||   :::|||:::|||||||||||||||||:::   :::      :::
 2099 ACTGCAGCGCCGGCATCGGCCGCACAGGCACCATCATTGTCATCGACATG 2148

  193 LeuLysAspAsn..................................... 196
      |||::::::|||
 2149 CTCATGGAGAACATCTCCACCAAGGGCCTGGACTGTGACATTGACATCCA 2198

  197 .......................ProHisSerAsnLeuGluGlnValA 205
                             ::::::|||::::::   |||:::
 2199 GAAGACCATCCAGATGGTGCGGGCGCAGCGCTCGGGCATGGTGCAGACGG 2248

  205 rgAlaAspPheArg...IleHisGlyThrIleGluCysTrpLysThrPro 220
        |||:::::::::   |||:::   :::|||
 2249 AGGCGCAGTACAAGTTCATCTACGTGGCCATC.................. 2280

  221 AlaGlnPheValGln.............LeuLysAlaMetGlnAlaGl 232
      |||||||||::::::           |||   ::::::|||:::||
 2281 GCCCAGTTCATTGAAACCACTAAGAAGAAGCTGGAGGTCCTGCAGTCGCA 2330

  232 n 232
      |
 2331 G 2331
```

seq_name: N_Geneseq_36:V71076

seq_documentation_block:
ID V71076 standard; cDNA; 2562 BP.
AC V71076;
DI 08-FEB-1999 (first entry)
DE Protein tyrosine phosphatase 1C-green fluorescent protein fusion product.
KW Human; protein tyrosine phosphatase 1C gene; fusion protein;
KW green fluorescent protein; GFP; intracellular signalling; chimera; ss.
OS Chimeric - Aequorea victoria.
OS Chimeric - Homo sapiens.
FH Key Location/Qualifiers
FT CDS 1. .2562
FT /*tag= a
PN WO9845704-A2.
PD 15-OCT-1998.
PF 07-APR-1998; DK0145.
PR 07-APR-1997; DK-000392.
PA (NOVO) NOVO-NORDISK AS.
PI Kasper A, Petersen Bjorn S, Scudder K, Thastrup O,
PI Tullin S;
DR WPI; 98-594491/50.
DR P-PSDB; W85031.
PT Determining effect on signalling pathways in live cells from
PT redistribution of luminophores - specifically fusions of green
PT fluorescent protein with a signalling component, and new apparatus,
PT particularly for identifying toxins and potential therapeutic agents
PS Claim 63; Pages 220-223; 326pp; English.
CC The present sequence encodes a human protein tyrosine phosphatase
CC 1C-green fluorescent protein (GFP) fusion protein. The fusion protein is
CC used in an assay to exemplify the invention. The specification describes
CC how quantitative information about the influence of a molecule on a
CC cellular response is obtained by recording the variation, caused by
CC the molecule, on mechanically intact living cells, in the spatially
CC distributed light emitted from a luminophore present in the cells. The
CC variation in light emission is processed to provide information that
CC correlates spatial distribution to the degree of the molecule. The method
CC is used to identify agents that (in)directly affect intracellular
CC signalling, especially to screen for potential therapeutic agents or
CC toxins, and to identify new drug targets.
SQ Sequence 2562 BP; 617 A; 756 C; 767 G; 422 T;

alignment_scores:
Quality: 82.50 Length: 167
Ratio: 1.058 Gaps: 10
Percent Similarity: 46.707 Percent Identity: 22.156

alignment_block:
US-09-068-804-12 x V71076 ..

Align seg 1/1 to: V71076 from: 1 to: 2562

```
  119 SerCysGlyGlu.........LysArgTyrThrIleProValLeuHisVa 132
      :::|||||||||         :::   |||:::::::   :::|||:::||
 1147 AACTGCGGGGAGCATGACACAACCGAATACAAACTCCGTACCTTACAGGT 1196

  132 lLys..................................................... 133
      |
 1197 CTCCCCGCTGGACAATGGAGACCTGATTCGGGAGATCTGGCATTACCAGT 1246

  134 .....AsnTrpProAspHis...GlnProLeuProSerThrAspGlnLeu 147
           :::|||||||||||||      |||      :::   :::   |||
 1247 ACCTGAGCTGGCCCGACCATGGGGTCCCCAGTGAGCCTGGGGGTGTCCTC 1296

  148 GluTyrLeuAlaAspArgValLysAsnSerAsnGlnAsnGlyAlaProGl 164
        :::|||   |||:::::::::::   :::::::::
 1297 AGCTTCCTG...GACCAGATCAACCAGCGGCAGGAAAGT.......... 1332

  164 yArgSerSerSerAspLysHisLeuPro.............MetIleH 176
                          ||||||               ::::::|
 1333 ....................CTGCCTCACGCAGGGCCCATCATCGTGC 1360

  176 isCysLeuGlyGlyValGlyArgThrGlyThrMetAlaAlaAlaLeuVal 192
      |||||   :::|||:::|||||||||||||||||:::   :::      :::
 1361 ACTGCAGCGCCGGCATCGGCCGCACAGGCACCATCATTGTCATCGACATG 1410

  193 LeuLysAspAsn..................................... 196
      |||::::::|||
 1411 CTCATGGAGAACATCTCCACCAAGGGCCTGGACTGTGACATTGACATCCA 1460

  197 .......................ProHisSerAsnLeuGluGlnValA 205
                             ::::::|||::::::   |||:::
 1461 GAAGACCATCCAGATGGTGCGGGCGCAGCGCTCGGGCATGGTGCAGACGG 1510

  205 rgAlaAspPheArg...IleHisGlyThrIleGluCysTrpLysThrPro 220
        |||:::::::::   |||:::   :::|||
 1511 AGGCGCAGTACAAGTTCATCTACGTGGCCATC.................. 1542

  221 AlaGlnPheValGln.............LeuLysAlaMetGlnAlaGl 232
      |||||||||::::::           |||   ::::::|||:::||
 1543 GCCCAGTTCATTGAAACCACTAAGAAGAAGCTGGAGGTCCTGCAGTCGCA 1592

  232 n 232
      |
 1593 G 1593
```

seq_name: N_Geneseq_36:V04868

seq_documentation_block:
ID V04868 standard; DNA; 4752 BP.
AC V04868;
DT 30-APR-1998 (first entry)
DE Nucleotide sequence of human glomerular epithelial protein 1.
KW Glomerular epithelial protein 1; GLEPP1; podocyte; gene therapy;
KW fibronectin type III-like repeat; protein tyrosine phosphatase domain;
KW PTPase domain; glomerular disorder; low level; GLEPP1 expression; ds.
OS Homo sapiens.

```
FH   Key             Location/Qualifiers
FT   CDS             808. .4278
FT                   /*tag=  a
PN   US5705623-A.
PD   06-JAN-1998.
PF   25-FEB-1994; 201697.
PR   25-FEB-1994; US-201697.
PA   (UNMI ) UNIV MICHIGAN.
PI   Thomas PE, Wiggins RC;
DR   WPI; 98-086195/08.
DR   P-PSDB; W42991.
PT   DNA encoding glomerular epithelial protein 1 - useful for gene
PT   therapy of glomerular disorders
PS   Claim 3; Fig 5A-C; 47pp; English.
CC   The present sequence encodes a human glomerular epithelial protein 1
CC   (GLEPP1). This protein is primarily expressed in mammalian podocytes.
CC   The protein has a large extracellular domain with a sequence containing
CC   8 fibronectin type III-like repeats, a hydrophobic transmembrane segment
CC   and a single intracellular protein tyrosine phosphatase (PTPase) domain.
CC   The present sequence was isolated from a human kidney cDNA library
CC   DNA, which was probed with rabbit GLEPP1 probes. Three different
CC   GLEPP1 proteins were identified, the full length protein (present
CC   sequence), an alternatively spliced version (V04886), and a truncated
CC   version (V04887). The nucleic acids are used for gene therapy of
CC   glomerular disorders associated with abnormally low levels of GLEPP1
CC   expression.
SQ   Sequence   4752 BP;    1322 A;    1093 C;    1075 G;    1262 T;


alignment_scores:
            Quality:   84.50            Length:    198
              Ratio:  0.775              Gaps:     10
 Percent Similarity:  55.051   Percent Identity: 20.202

alignment_block:
 US-09-068-804-12 x V04868   ..

 Align seg 1/1  to: V04868  from: 1  to: 4752

      43 IleArgThrGlnLysIleArgProAspArgLeuGluAlaHisMetLysMe 59
         |||   ||||||         |||:::      ::::::      |||||
    3745 ATTGCCACCCAGGGGCCACTGCCTGAAACCAGAAATGACTTCTGGAAGAT 3794

      59 tLeuLeuGluLysGluCysSerCysLeuValValLeuThrSerGluArgS 76
         |:::|||::::::   :::      :::|||:::||||||      ::::
    3795 GGTCCTGCAACAAAAGTCTCAGATTATTGTCATGCTCACT......CAGT 3838

      76 erAspAlaGlyLysThrIleThr..............ThrTyrPheArg 87
         ::::::::   :::   ::::::              ::::::
    3839 GTAATGAGAAAAGGAGGGTGAAATGTGACCATTACTGGCCATTCACGGAA 3888

      88 GlySerTyrThrPheGlyGluValHisThrAsnSerGlnLysValSerSe 104
            :::   :::::::|||::::::      :::   ::::::::|||
    3889 GAACCTATAGCCTATGGAGACATC......ACTGTGGAGATGATTTCAGA 3932

     104 rAlaSerGlnGlyGlu...AlaIleAspGlnTyrAsnMetGlnLeuSerC 120
          :::   |||:::::   |||      ::::::      ::::::::::
    3933 GGAAGAGCAGGACGACTGGGCCTGTAGACACTTC......CGGATCAACT 3976

     120 ysGlyGluLysArgTyrThrIleProValLeuHisValLysAsnTrpPro 136
         ::::::::          :::      ::::::   ::::::||||||
    3977 ATGCTGACGAGATGCAGGATGTGATGCATTTTAACTACACTGCATGGCCT 4026

     137 AspHisGlnProLeuProSerThrAspGlnLeuGluTyrLeuAlaAspAr 153
         ||||||      :::|||::::::::      |||   :::   :::
    4027 GATCAT...GGTGTGCCCACAGCAAATGCTGCAGAAAGTATCCTGCAGTT 4073

     153 gValLysAsnSerAsnGlnAsnGlyAlaProGlyArgSerSerSerAspL 170
          |||           |||:::::::::   :::::::::
    4074 TGTACACATGGTCCGACAGCAAGCTACCAAGAGCAAAGGT.......... 4113

     170 ysHisLeuProMet...IleHisCysLeuGlyGlyValGlyArgThrGly 185
                 ||||||   |||||||||   :::|||||||||||||||||||
```

```
    4114 ........CCCATGATCATTCACTGCAGTGCTGGCGTGGGACGGACAGGA 4155

     186 ThrMetAlaAlaAlaLeuValLeuLysAspAsnProHisSerAsnLeuGl 202
         |||      |||                                 |||::
    4156 ACATTCATTGCC.................................CTGGA 4172

     202 uGlnValArgAlaAspPheArgIleHisGlyThrIleGluCysTrpLysT 219
         :::::::      :::::::|||   |||
    4173 CAGGCTCTTGCAGCACATTCGGGATCAT..................... 4200

     219 hrProAlaGlnPheValGlnLeuLysAlaMetGlnAlaGlnLeu 233
                :::||||||::::::   ::::::   :::::::::
    4201 ........GAGTTTGTTGACATCTTAGGGCTGGTGTCAGAAATG 4236

seq_name: N_Geneseq_36:V04887

seq_documentation_block:
ID   V04887 standard; DNA; 4799 BP.
AC   V04887;
DT   30-APR-1998  (first entry)
DE   Nucleotide sequence of a human glomerular epithelial protein 1.
KW   Glomerular epithelial protein 1; GLEPP1; podocyte; gene therapy;
KW   fibronectin type III-like repeat; protein tyrosine phosphatase domain;
KW   PTPase domain; glomerular disorder; low level; GLEPP1 expression; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             808. .3606
FT                   /*tag=  a
FT   Misc_fetaure    3445. .3497
FT                   /*tag=  b
FT                   note= "additional nucleotides, as compared to V04868"
PN   US5705623-A.
PD   06-JAN-1998.
PF   25-FEB-1994; 201697.
PR   25-FEB-1994; US-201697.
PA   (UNMI ) UNIV MICHIGAN.
PI   Thomas PE, Wiggins RC;
DR   WPI; 98-086195/08.
PT   DNA encoding glomerular epithelial protein 1 - useful for gene
PT   therapy of glomerular disorders
PS   Claim 6; Fig 7A-B; 47pp; English.
CC   The present sequence encodes a human glomerular epithelial protein 1
CC   (GLEPP1). This protein is primarily expressed in mammalian podocytes.
CC   The protein has a large extracellular domain with a sequence containing
CC   8 fibronectin type III-like repeats and a hydrophobic transmembrane
CC   segment. The present sequence was isolated from a human kidney cDNA
CC   library DNA, which was probed with rabbit GLEPP1 probes. Three different
CC   GLEPP1 proteins were identified, the full length protein (V04868),
CC   an alternatively spliced version (V04886), and a truncated version
CC   (present sequence). The nucleic acids are used for gene therapy of
CC   glomerular disorders associated with abnormally low levels of GLEPP1
CC   expression.
SQ   Sequence   4799 BP;    1340 A;    1100 C;    1084 G;    1275 T;


alignment_scores:
            Quality:   84.50            Length:    198
              Ratio:  0.775              Gaps:     10
 Percent Similarity:  55.051   Percent Identity: 20.202

alignment_block:
 US-09-068-804-12 x V04887   ..

 Align seg 1/1  to: V04887  from: 1  to: 4799

      43 IleArgThrGlnLysIleArgProAspArgLeuGluAlaHisMetLysMe 59
         |||   ||||||         |||:::      ::::::      |||||
    3792 ATTGCCACCCAGGGGCCACTGCCTGAAACCAGAAATGACTTCTGGAAGAT 3841

      59 tLeuLeuGluLysGluCysSerCysLeuValValLeuThrSerGluArgS 76
         |:::|||::::::   :::      :::|||:::||||||      ::::
    3842 GGTCCTGCAACAAAAGTCTCAGATTATTGTCATGCTCACT......CAGT 3885
```

```
     76 erAspAlaGlyLysThrIleThr.............ThrTyrPheArg 87
        ::::::::   :::   ::::::             ::::::
   3886 GTAATGAGAAAAGGAGGGTGAAATGTGACCATTACTGGCCATTCACGGAA 3935

     88 GlySerTyrThrPheGlyGluValHisThrAsnSerGlnLysValSerSe 104
          :::   ::::::|||::::::      :::   :::::::::|||
   3936 GAACCTATAGCCTATGGAGACATC......ACTGTGGAGATGATTTCAGA 3979

    104 rAlaSerGlnGlyGlu...AlaIleAspGlnTyrAsnMetGlnLeuSerC 120
         :::   |||::::::   |||      ::::::      ::::::::::
   3980 GGAAGAGCAGGACGACTGGGCCTGTAGACACTTC......CGGATCAACT 4023

    120 ysGlyGluLysArgTyrThrIleProValLeuHisValLysAsnTrpPro 136
        ::::::::            :::      ::::::   ::::::||||||
   4024 ATGCTGACGAGATGCAGGATGTGATGCATTTTAACTACACTGCATGGCCT 4073

    137 AspHisGlnProLeuProSerThrAspGlnLeuGluTyrLeuAlaAspAr 153
        ||||||      :::|||:::::::::      |||   :::   :::
   4074 GATCAT...GGTGTGCCCACAGCAAATGCTGCAGAAAGTATCCTGCAGTT 4120

    153 gValLysAsnSerAsnGlnAsnGlyAlaProGlyArgSerSerSerAspL 170
         |||            |||:::::::::   :::::::::
   4121 TGTACACATGGTCCGACAGCAAGCTACCAAGAGCAAAGGT.......... 4160

    170 ysHisLeuProMet...IleHisCysLeuGlyGlyValGlyArgThrGly 185
                |||||||   |||||||||   :::||||||||||||||||||
   4161 ........CCCATGATCATTCACTGCAGTGCTGGCGTGGGACGGACAGGA 4202

    186 ThrMetAlaAlaAlaLeuValLeuLysAspAsnProHisSerAsnLeuGl 202
        |||       |||                                 |||::
   4203 ACATTCATTGCC.................................CTGGA 4219

    202 uGlnValArgAlaAspPheArgIleHisGlyThrIleGluCysTrpLysT 219
        :::::::      :::::::|||   |||
   4220 CAGGCTCTTGCAGCACATTCGGGATCAT..................... 4247

    219 hrProAlaGlnPheValGlnLeuLysAlaMetGlnAlaGlnLeu 233
               :::|||||||::::::   ::::::   :::::::::
   4248 ........GAGTTTGTTGACATCTTAGGGCTGGTGTCAGAAATG 4283

seq_name: N_Geneseq_36:V04886

seq_documentation_block:
ID   V04886 standard; DNA; 4815 BP.
AC   V04886;
DT   30-APR-1998  (first entry)
DE   Nucleotide sequence of a human glomerular epithelial protein 1.
KW   Glomerular epithelial protein 1; GLEPP1; podocyte; gene therapy;
KW   fibronectin type III-like repeat; protein tyrosine phosphatase domain;
KW   PTPase domain; glomerular disorder; low level; GLEPP1 expression; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             808. .4437
FT                   /*tag=  a
PN   US5705623-A.
PD   06-JAN-1998.
PF   25-FEB-1994; 201697.
PR   25-FEB-1994; US-201697.
PA   (UNMI ) UNIV MICHIGAN.
PI   Thomas PE, Wiggins RC;
DR   WPI; 98-086195/08.
PT   DNA encoding glomerular epithelial protein 1 - useful for gene
PT   therapy of glomerular disorders
PS   Claim 5; Fig 6A-B; 47pp; English.
CC   The present sequence encodes a human glomerular epithelial protein 1
CC   (GLEPP1). This protein is primarily expressed in mammalian podocytes.
CC   The protein has a large extracellular domain with a sequence containing
CC   8 fibronectin type III-like repeats, a hydrophobic transmembrane segment
CC   and a single intracellular protein tyrosine phosphatase (PTPase) domain.
CC   The present sequence was isolated from a human kidney cDNA library
CC   DNA, which was probed with rabbit GLEPP1 probes. Three different
CC   GLEPP1 proteins were identified, the full length protein (V04868),
CC   an alternatively spliced version (present sequence), and a truncated
CC   version (V04887). The nucleic acids are used for gene therapy of
CC   glomerular disorders associated with abnormally low levels of GLEPP1
CC   expression.
SQ   Sequence   4815 BP;    1335 A;    1107 C;    1094 G;    1279 T;


alignment_scores:
            Quality:   84.50              Length:    198
              Ratio:  0.775                Gaps:     10
 Percent Similarity:  55.051   Percent Identity: 20.202

alignment_block:
 US-09-068-804-12 x V04886   ..

 Align seg 1/1  to: V04886  from: 1  to: 4815

     43 IleArgThrGlnLysIleArgProAspArgLeuGluAlaHisMetLysMe 59
        |||   ||||||          |||:::      ::::::      |||||
   3808 ATTGCCACCCAGGGGCCACTGCCTGAAACCAGAAATGACTTCTGGAAGAT 3857

     59 tLeuLeuGluLysGluCysSerCysLeuValValLeuThrSerGluArgS 76
        |:::|||::::::   :::      :::|||:::||||||      ::::
   3858 GGTCCTGCAACAAAAGTCTCAGATTATTGTCATGCTCACT......CAGT 3901

     76 erAspAlaGlyLysThrIleThr.............ThrTyrPheArg 87
        ::::::::   :::   ::::::             ::::::
   3902 GTAATGAGAAAAGGAGGGTGAAATGTGACCATTACTGGCCATTCACGGAA 3951

     88 GlySerTyrThrPheGlyGluValHisThrAsnSerGlnLysValSerSe 104
          :::   ::::::|||::::::      :::   :::::::::|||
   3952 GAACCTATAGCCTATGGAGACATC......ACTGTGGAGATGATTTCAGA 3995

    104 rAlaSerGlnGlyGlu...AlaIleAspGlnTyrAsnMetGlnLeuSerC 120
         :::   |||:::::::   |||      ::::::      ::::::::::
   3996 GGAAGAGCAGGACGACTGGGCCTGTAGACACTTC......CGGATCAACT 4039

    120 ysGlyGluLysArgTyrThrIleProValLeuHisValLysAsnTrpPro 136
        ::::::::            :::      ::::::   ::::::||||||
   4040 ATGCTGACGAGATGCAGGATGTGATGCATTTTAACTACACTGCATGGCCT 4089

    137 AspHisGlnProLeuProSerThrAspGlnLeuGluTyrLeuAlaAspAr 153
        ||||||      :::|||:::::::::      |||   :::   :::
   4090 GATCAT...GGTGTGCCCACAGCAAATGCTGCAGAAAGTATCCTGCAGTT 4136

    153 gValLysAsnSerAsnGlnAsnGlyAlaProGlyArgSerSerSerAspL 170
         |||            |||:::::::::   :::::::::
   4137 TGTACACATGGTCCGACAGCAAGCTACCAAGAGCAAAGGT.......... 4176

    170 ysHisLeuProMet...IleHisCysLeuGlyGlyValGlyArgThrGly 185
                |||||||   |||||||||   :::||||||||||||||||||
   4177 ........CCCATGATCATTCACTGCAGTGCTGGCGTGGGACGGACAGGA 4218

    186 ThrMetAlaAlaAlaLeuValLeuLysAspAsnProHisSerAsnLeuGl 202
        |||       |||                                 |||::
   4219 ACATTCATTGCC.................................CTGGA 4235

    202 uGlnValArgAlaAspPheArgIleHisGlyThrIleGluCysTrpLysT 219
        :::::::      :::::::|||   |||
   4236 CAGGCTCTTGCAGCACATTCGGGATCAT..................... 4263

    219 hrProAlaGlnPheValGlnLeuLysAlaMetGlnAlaGlnLeu 233
               :::|||||||::::::   ::::::   :::::::::
   4264 ........GAGTTTGTTGACATCTTAGGGCTGGTGTCAGAAATG 4299

seq_name: N_Geneseq_36:T48479

seq_documentation_block:
ID   T48479 standard; cDNA to mRNA; 5113 BP.
AC   T48479;
DT   21-MAR-1997  (first entry)
DE   Human tyrosine dephosphorylase coding sequence.
KW   Tyrosine dephosphorylase; drug development; differentiation control;
KW   cancer; neoplasia; proliferative disease; tyrosine phosphatase; ds.
```

```
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             121. .3771
FT                   /*tag= a
PN   J08242852-A.
PD   24-SEP-1996.
PF   10-MAR-1995; 051374.
PR   10-MAR-1995; JP-051374.
PA   (MITU ) MITSUBISHI CHEM CORP.
DR   WPI; 96-479899/48.
DR   P-PSDB; W10685.
PT   Tyrosine dephosphorylase and related gene - useful for development
PT   of drugs controlling differentiation and growth of cells
PS   Claim 1; Page 7-12; 14pp; Japanese.
CC   T48479 is a cDNA sequence encoding human tyrosine dephosphorylase
CC   designated PTP-U2. The PTP-U2 gene and protein are useful for the
CC   development of drugs that control differentiation and cell growth.
CC   Such drugs are useful for treating proliferative diseases e.g.
CC   cancer.
SQ   Sequence   5113 BP;    1442 A;    1111 C;    1114 G;    1446 T;


alignment_scores:
           Quality:   84.50            Length:    198
             Ratio:  0.775              Gaps:     10
 Percent Similarity:  55.051   Percent Identity: 20.202

alignment_block:
 US-09-068-804-12 x T48479   ..

 Align seg 1/1  to: T48479  from: 1  to: 5113

      43 IleArgThrGlnLysIleArgProAspArgLeuGluAlaHisMetLysMe 59
         |||   ||||||         |||:::      ::::::      |||||
    3142 ATTGCCACCCAGGGGCCACTGCCTGAAACCAGAAATGACTTCTGGAAGAT 3191

      59 tLeuLeuGluLysGluCysSerCysLeuValValLeuThrSerGluArgS 76
         |:::|||:::::   :::     :::|||:::||||||      ::::
    3192 GGTCCTGCAACAAAAGTCTCAGATTATTGTCATGCTCACT......CAGT 3235

      76 erAspAlaGlyLysThrIleThr.............ThrTyrPheArg 87
         ::::::::   :::   ::::::               ::::::
    3236 GTAATGAGAAAAGGAGGGTGAAATGTGACCATTACTGGCCATTCACGGAA 3285

      88 GlySerTyrThrPheGlyGluValHisThrAsnSerGlnLysValSerSe 104
            :::   ::::::|||:::::::      :::   ::::::::::|||
    3286 GAACCTATAGCCTATGGAGACATC......ACTGTGGAGATGATTTCAGA 3329

     104 rAlaSerGlnGlyGlu...AlaIleAspGlnTyrAsnMetGlnLeuSerC 120
          :::   |||:::::   |||      ::::::      ::::::::::
    3330 GGAAGAGCAGGACGACTGGGCCTGTAGACACTTC......CGGATCAACT 3373

     120 ysGlyGluLysArgTyrThrIleProValLeuHisValLysAsnTrpPro 136
         ::::::::            :::      ::::::   ::::::||||||
    3374 ATGCTGACGAGATGCAGGATGTGATGCATTTTAACTACACTGCATGGCCT 3423

     137 AspHisGlnProLeuProSerThrAspGlnLeuGluTyrLeuAlaAspAr 153
         ||||||      :::|||::::::::::      |||   :::   :::
    3424 GATCAT...GGTGTGCCCACAGCAAATGCTGCAGAAAGTATCCTGCAGTT 3470

     153 gValLysAsnSerAsnGlnAsnGlyAlaProGlyArgSerSerSerAspL 170
          |||           |||::::::::::   :::::::::
    3471 TGTACACATGGTCCGACAGCAAGCTACCAAGAGCAAAGGT.......... 3510

     170 ysHisLeuProMet...IleHisCysLeuGlyGlyValGlyArgThrGly 185
               ||||||   |||||||||   :::||||||||||||||||||
    3511 ........CCCATGATCATTCACTGCAGTGCTGGCGTGGGACGGACAGGA 3552

     186 ThrMetAlaAlaAlaLeuValLeuLysAspAsnProHisSerAsnLeuGl 202
         |||      |||                                 |||::
    3553 ACATTCATTGCC.................................CTGGA 3569

     202 uGlnValArgAlaAspPheArgIleHisGlyThrIleGluCysTrpLysT 219
```

```
         :::::::      :::::::|||   |||
    3570 CAGGCTCTTGCAGCACATTCGGGATCAT..................... 3597

     219 hrProAlaGlnPheValGlnLeuLysAlaMetGlnAlaGlnLeu 233
               :::|||||||::::::   ::::::   :::::::::
    3598 ........GAGTTTGTTGACATCTTAGGGCTGGTGTCAGAAATG 3633

seq_name: N_Geneseq_36:Q21001

seq_documentation_block:
ID   Q21001 standard; DNA; 2872 BP.
AC   Q21001;
DT   28-MAY-1992  (first entry)
DE   Murine receptor-type protein tyrosine phosphatase gene.
KW   R-PTPase; cellular metabolism; cancer; diabetes; mouse; ss.
OS   Mus musculus.
FH   Key             Location/Qualifiers
FT   cds             259. .2640
FT                   /*tag= a
PN   WO9201050-A.
PD   23-JAN-1992.
PF   11-JUL-1991; U04892.
PR   11-JUL-1990; US-551270.
PR   26-FEB-1991; US-654188.
PA   (UYNY-) NEW YORK UNIV.
PI   Schlessinger J;
DR   WPI; 92-056865/07.
DR   P-PSDB; R20743.
PT   Human receptor-type protein tyrosine phosphatase - has DNA
PT   encoding it and antibodies specific for it, useful for screening
PT   drugs affecting R-PTPase activity, and detection of mutant genes
PS   Claim 7; Fig 1; 77pp; English.
CC   The sequence is that of the gene encoding mouse receptor-type
CC   protein tyrosine phosphatase (R-PTPase). It may be used to produce
CC   the R-PTPase recombinantly. Probes may be designed from the gene
CC   that can be used to detect mutant genes in PCR amplification and
CC   hybridisation assays.
SQ   Sequence   2872 BP;    785 A;    721 C;    724 G;    642 T;


alignment_scores:
           Quality:   81.50            Length:    174
             Ratio:  0.832              Gaps:      8
 Percent Similarity:  56.322   Percent Identity: 22.989

alignment_block:
 US-09-068-804-12 x Q21001   ..

 Align seg 1/1  to: Q21001  from: 1  to: 2872

      43 IleArgThrGlnLysIleArgProAspArgLeuGluAlaHisMetLysMe 59
         |||   :::|||      :::   :::   :::::::::      :::||
    1162 ATCGCTGCACAAGGACCAAAAGAAGAAACAGTGAATGACTTCTGGAGAAT 1211

      59 tLeuLeuGluLysGluCysSerCysLeuValValLeuThrSer......G 74
         |:::   |||::::::   :::   :::|||::::::|||:::      |
    1212 GATATGGGAACAAAACACAGCTACTATTGTCATGGTGACCAACCTGAAGG 1261

      74 luArgSerAspAlaGlyLysThrIleThrThrTyrPhe......ArgGly 88
         |||||   :::          :::   :::   |||          :::|||
    1262 AGAGAAAGGAGTGT......AAATGTGCCCAATACTGGCCAGACCAAGGC 1305

      89 SerTyrThrPheGlyGluValHisThrAsnSerGlnLysValSerSerAl 105
         :::   |||:::|||:::|||:::::::::   :::   |||:::
    1306 TGCTGGACCTATGGGAATGTCCGTGTGTCTGTCGAGGATGTGACTGTTCT 1355

     105 aSerGln.................................GlyGluAlaI 111
              :::                               |||:::::::
    1356 GGTGGACTACACAGTACGGAAATTCTCGATCCAGCAGGTGGGCGACGTGA 1405

     111 leAspGlnTyrAsnMetGlnLeuSerCysGlyGluLysArgTyrThrIle 127
         ::::::::         :::|||                         |||
    1406 CCAACAGGAAACCACAGCGCCTC.......................ATC 1431
```

```
 128 ProValLeuHisValLysAsnTrpProAspHis...GlnProLeuProSe 143
     :::   :::|||   ::::::||||||||||         |||::::::::
1432 ACTCAGTTCCACTTCACCAGCTGGCCAGACTTTGGGGTGCCTTTCACCCC 1481

 143 rThrAspGlnLeuGluTyrLeuAlaAspArgValLysAsnSerAsnGlnA 160
     :::::::   |||   :::|||       :::||||||::::::|||:::: 
1482 AATTGGCATGCTCAAGTTCCTC...AAGAAGGTGAAGGCCTGTAACCCTC 1528

 160 snGlyAlaProGlyArgSerSerSerAspLysHisLeuProMetIleHis 176
     ::   |||   :::                          ::::::|||
1529 AGTACGCAGGGGCTATC.....................GTGGTCCAC 1554

 177 CysLeuGlyGlyValGlyArgThrGlyThrMetAlaAlaAlaLeuValLe 193
     |||   :::||||||||||||||||||||||||   ::::::      :::::
1555 TGCAGTGCAGGTGTAGGGCGCACTGGCACCTTTGTTGTCATCGATGCCAT 1604

 193 uLysAspAsnProHisSerAsn 200
     :   |||      ||||||:::
1605 GCTGGACATGATGCATTCGGAG 1626
```

This Page Blank (uspto)

```
OM of: US-09-068-804-12 to: GenEmbl:*   out_format : pfs

Date: Oct 5, 1999  8:00 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn2_1/USPTO_spool/US09068804/runat_05101999_145134_10315/app_query.fasta.1
 -DB=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=pam150 -TRANS=human40.cdi
 -LIST=45 -DOCALIGN=200 -THR_SCORE=escore -ALIGN=15 -MODE=LOCAL
 -OUTFMT=pfs -NORM=stat -USER=US09068804 -NCPU=6 -ICPU=3 -WAIT
 -THREADS=1

Search information block:
Query: US-09-068-804-12
Query length: 239
Database: GenEmbl:*
Database sequences: 679419
Database length: 1590154680
Search time (sec): 1424.920000


score_list:
Sequence              Strd Orig    ZScore    EScore Len  ! Documentation  ..
gb_ba2:STU63293         +  949.50 2090.99  1.6e-108  1691  ! U63293 Salmonella typhimurium
gb_ba1:STSIPSTP         +  941.50 2065.27  4.3e-107  3964  ! X92546 S.typhi sipA, sipF, and
gb_ba1:YEPYOP51A        +  202.50  422.67  1.3e-15   2000  ! M30457 Yersinia enterocolitica
gb_ba1:YPYOPH           +  202.50  419.16  2.1e-15   2924  ! Y00551 Yersinia pseudotubercul
gb_ba2:AF102990         +  202.50  389.86  9.0e-14   69673  ! AF102990 Yersinia enterocolit
gb_ba2:AF053946         +  202.50  389.75  9.1e-14   70504  ! AF053946 Yersinia pestis plas
gb_ba2:AF074612         +  202.50  389.74  9.1e-14   70559  ! AF074612 Yersinia pestis plas
gb_pat:A52315           -  123.00  243.58  1.3e-05   2405  ! A52315 Sequence 19 from Patent
gb_ba1:STIAGAB          -  123.00  243.58  1.3e-05   2406  ! X80892 S.typhi genes iagA and
gb_in1:DDIPTP1A         +  109.00  215.29  0.0005    1749  ! L07125 Dictyostelium discoideu
gb_in1:AF017083         +  102.50  189.05  0.0137    6227  ! AF017083 Hirudo medicinalis re
gb_pr2:HSRPTYRPH        +   99.00  184.47  0.0247    4389  ! X95712 H.sapiens mRNA for rece
gb_ro:RRU66566          +   91.00  175.79  0.0751    1627  ! U66566 Rattus rattus rat recep
gb_pr2:HSLARR           +   94.50  169.23  0.1741    7702  ! Y00815 Human mRNA for LCA-homo
gb_ov:XLU15287          +   89.50  168.79  0.1843    2416  ! U15287 Xenopus laevis phosphot
gb_ro:MMLARMRN          +   88.50  167.42  0.2198    2202  ! Z37988 M.musculus LAR mRNA enc
gb_ro:U00477            -   88.50  167.37  0.2211    2213  ! U00477 Rattus norvegicus LAR r
gb_pl2:AF055635         -   86.00  167.00  0.2319    1260  ! AF055635 Arabidopsis thaliana
gb_ro:RATOSTP           +   92.00  166.86  0.2361    5446  ! L36884 Rattus norvegicus prote
gb_in1:DRODPTP99A       +   90.50  166.16  0.2581    4087  ! M80539 D.melanogaster receptor
gb_pr2:HSU60289         +   91.00  165.91  0.2668    4742  ! U60289 Human receptor protein
gb_in1:DRODPTP          +   90.50  165.55  0.2793    4368  ! M81795 Drosophila melanogaster
gb_ro:RATPTPECA         +   87.50  165.39  0.2852    2155  ! D78610 Rat mRNA for protein ty
gb_ro:MMU40280          +   87.50  165.15  0.2938    2210  ! U40280 Mus musculus putative p
gb_ov:GGU65891          +   90.00  165.13  0.2949    4053  ! U65891 Gallus gallus protein t
gb_ro:MMU36488          +   91.00  164.98  0.3003    5240  ! U36488 Mus musculus embryonic
gb_pr3:HSU71075         +   91.00  164.82  0.3067    5333  ! U71075 Homo sapiens protein ty
gb_ro:MMU62387          +   87.50  164.42  0.3228    2392  ! U62387 Mus musculus protein ty
gb_pr2:HSU73727         +   91.00  164.39  0.3242    5589  ! U73727 Human receptor protein
gb_in1:CEF38A3          +   96.00  163.51  0.3629    20555  ! Z49938 Caenorhabditis elegans
gb_ro:MUSPTPE           +   87.50  162.88  0.3934    2827  ! D83484 Mouse mRNA for protein
gb_ro:RNLAR2            -   88.50  162.04  0.4382    3941  ! X83546 R.norvegicus mRNA for l
gb_in1:DRODPTP99B       -   90.50  161.65  0.4605    6660  ! M80464 D.melanogaster protein
gb_ro:M60103            +   88.50  161.47  0.4712    4190  ! M60103 Rattus norvegicus leuk
gb_ov:GGU38620          +   85.50  160.95  0.5041    2150  ! U38620 Gallus gallus protein t
gb_ro:RNLAR1            +   88.50  160.44  0.5379    4685  ! X83505 R.norvegicus mRNA for l
gb_ro:D88187            +   89.00  160.13  0.5594    5464  ! D88187 Mus musculus mRNA for F
gb_ro:MMU55057          +   89.00  159.69  0.5921    5732  ! U55057 Mus musculus receptor p
gb_ro:MMU37467          +   85.50  159.60  0.5988    2486  ! U37467 Mus musculus protein ty
gb_pr2:HSPTP1D          +   84.50  159.48  0.6087    1980  ! X70766 H.sapiens mRNA for phos
gb_ro:MMU37466          +   85.50  158.92  0.6534    2676  ! U37466 Mus musculus protein ty
gb_pr1:HUMPTPASA        +   84.50  158.84  0.6604    2121  ! L07527 Homo sapiens protein-ty
gb_ro:MMU37465          +   85.50  158.64  0.6778    2760  ! U37465 Mus musculus protein
gb_ro:RATBEM1           +   85.50  158.21  0.7155    2889  ! D45412 Rat mRNA for brain-e
gb_pat:I23561           +   84.50  158.19  0.7180    2276  ! I23561 Sequence 11 from pat

seq_name: gb_ba2:STU63293

seq_documentation_block:
LOCUS       STU63293     1691 bp    DNA             BCT       05-SEP-1996
DEFINITION  Salmonella typhimurium protein tyrosine phosphatase SptP (SptP)
            gene, complete cds.
ACCESSION   U63293
NID         g1519053
VERSION     U63293.1  GI:1519053
KEYWORDS    .
SOURCE      Salmonella typhimurium.
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Salmonella.
REFERENCE   1  (bases 1 to 1691)
  AUTHORS   Kaniga,K., Uralil,J., Bliska,J.B. and Galan,J.E.
  TITLE     A secreted protein tyrosine phosphatase with modular effector
            domains in the bacterial pathogen Salmonella typhimurium
  JOURNAL   Mol. Microbiol. 21 (3), 633-641 (1996)
  MEDLINE   97020031
REFERENCE   2  (bases 1 to 1691)
  AUTHORS   Galan,J.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUL-1996) Molecular Genetics and Microbiology, State
            University of New York at Stony Brook, Stony Brook, NY 11794-522,
            USA
FEATURES             Location/Qualifiers
     source          1. .1691
                     /organism="Salmonella typhimurium"
                     /strain="SL1344"
                     /db_xref="taxon:602"
     gene            11. .1642
                     /gene="SptP"
     CDS             11. .1642
                     /gene="SptP"
                     /note="involved in the virulence of Salmonella
                     typhimurium"
                     /codon_start=1
                     /transl_table=11
                     /product="protein tyrosine phosphatase SptP"
                     /protein_id="AAC44349.1"
                     /db_xref="PID:g1519054"
                     /db_xref="GI:1519054"
                     /translation="MLKYEERKLNNLTLSSFSKVGVSNDARLYIAKENTDKAYVAPEK
                     FSSKVLTWLGKMPLFKNTEVVQKHTENIRVQDQKILQTFLHALTEKYGETAVNDALLM
                     SRINMNKPLTQRLAVQITECVKAADEGFINLIKSKDNVGVRNAALVIKGGDTKVAEKN
                     NDVGAESKQPLLDIALKGLKRTLPQLEQMDGNSLRENFQEMASGNGPLRSLMTNLQNL
                     NKIPEAKQLNDYVTTLTNIQVGVARFSQWGTCGGEVERWVDKASTHELTQAVKKIHVI
                     AKELKNVTAELEKIEAGAPMPQTMSGPILGLARFAVSSIPINQQTQVKLSDGMPVPVN
                     TLTFDGKPVALAGSYPKNTPDALEAHMKMLLEKECSCLVVLTSEDQMQAKQLPPYFRG
                     SYTFGEVHTNSQKVSSASQGEAIDQYNMQLSCGEKRYIIPVLHVKNWPDHQPLPSTDQ
                     LEYLADRVKNSNQNGAPGRSSSDKHLPMIHCLGGVGRTGTMAAALVLKDNPHSNLEQV
                     RADFRDSRNNRMLEDASQFVQLKAMQAQLLMTTAS"
BASE COUNT       562 a    326 c    412 g    391 t
ORIGIN


alignment_scores:
            Quality:  949.50              Length:    241
              Ratio:  4.146                 Gaps:      4
 Percent Similarity:  95.021   Percent Identity: 94.606

alignment_block:
 US-09-068-804-12 x STU63293   ..

 Align seq 1/1  to: STU63293  from: 1  to: 1691

       2 ProGlyThrIleCysGlyGlnGlnHisSer.IleAsnGlnGlnThrGlnV 18
         |||||||||||||||||||||||||||||| ||||||||||||||||||
```

```
  928 CCTGGCACGATTTGCGGTCAGCAGCATTCCAATTAATCAGCAAACCCAGG 977

   18 alLysLeuSerAspGlyMetProValProValIleArgLeuThrPheAsp 34
      ||||||||||||||||||||||||||||||||      ||||||||||||
  978 TGAAACTGAGCGATGGAATGCCTGTGCCAGTGAATACGTTAACTTTTGAC 1027

   35 GlyLysProValAlaLeuAlaGlyIleArgThrGlnLysIleArgProAs 51
      ||||||||||||||||||||||||          :::|||      |||||
 1028 GGTAAGCCTGTGGCATTAGCCGGT...TCGTACCCAAAAAATACGCCGGA 1074

   51 pArgLeuGluAlaHisMetLysMetLeuLeuGluLysGluCysSerCysL 68
      |  ||||||||||||||||||||||||||||||||||||||||||||||||
 1075 TGCGCTGGAAGCGCATATGAAAATGCTTCTTGAAAAAGAATGCTCGTGCC 1124

   68 euValValLeuThrSerGluArgSerAspAlaGlyLysThrIleThrThr 84
      ||||||||||||||||   ||||||||||||||||||||||||||||||
 1125 TGGTGGTGTTAACGTCGGA.AGATCAGATGCAGGCAAAACAATTACCACC 1173

   85 .TyrPheArgGlySerTyrThrPheGlyGluValHisThrAsnSerGlnL 101
       ||||||||||||||||||||||||||||||||||||||||||||||||||
 1174 CTATTTCAGAGGGAGCTACACCTTTGGCGAGGTGCATACCAACAGCCAAA 1223

  101 ysValSerSerAlaSerGlnGlyGluAlaIleAspGlnTyrAsnMetGln 117
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1224 AAGTGAGCTCAGCCAGTCAGGGAGAAGCGATAGACCAATACAATATGCAA 1273

  118 LeuSerCysGlyGluLysArgTyrThrIleProValLeuHisValLysAs 134
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1274 CTGTCCTGCGGGGAAAAGCGGTATACAATCCCGGTATTGCATGTGAAAAA 1323

  134 nTrpProAspHisGlnProLeuProSerThrAspGlnLeuGluTyrLeuA 151
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1324 TTGGCCAGATCACCAGCCGTTACCGTCTACGGATCAGTTAGAATACCTGG 1373

  151 laAspArgValLysAsnSerAsnGlnAsnGlyAlaProGlyArgSerSer 167
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1374 CGGATAGGGTGAAAAATAGTAACCAAAATGGCGCGCCGGGGCGTAGTTCA 1423

  168 SerAspLysHisLeuProMetIleHisCysLeuGlyGlyValGlyArgTh 184
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1424 TCAGATAAGCATTTACCGATGATTCATTGTCTGGGCGGAGTGGGAAGAAC 1473

  184 rGlyThrMetAlaAlaAlaLeuValLeuLysAspAsnProHisSerAsnL 201
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1474 CGGAACGATGGCGGCGGCCCTTGTACTTAAGGATAATCCTCATAGTAATC 1523

  201 euGluGlnValArgAlaAspPheArg.IleHisGlyThrIleGluCysTr 217
      |||||||||||||||||||||||| ||||||||||||||||||||||||
 1524 TGGAGCAGGTACGTGCAGATTTCCGGGATTCACGGAACAATAGAATGCTG 1573

  217 pLysThrProAlaGlnPheValGlnLeuLysAlaMetGlnAlaGlnLeuL 234
      ||||||||||   ||||||||||||||||||||||||||||||||||||
 1574 GAAGACGCCTCA.CAGTTTGTACAACTAAAAGCAATGCAAGCCCAGTTGC 1622

  234 euMetThrThrAlaSer 239
      |||||||||||||||||
 1623 TTATGACGACGGCAAGC 1639

seq_name: gb_ba1:STSIPSTP

seq_documentation_block:
LOCUS       STSIPSTP     3964 bp    DNA             BCT       28-JAN-1997
DEFINITION  S.typhi sipA, sipF, and stpA genes.
ACCESSION   X92546
NID         g1514650
VERSION     X92546.1  GI:1514650
KEYWORDS    sipA gene; sipF gene; stpA gene.
SOURCE      Salmonella typhi.
  ORGANISM  Salmonella typhi
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Salmonella.
REFERENCE   1  (bases 1 to 3964)
  AUTHORS   Arricau,N., Hermant,D., Waxin,H. and Popoff,M.Y.
  TITLE     Molecular characterization of the Salmonella typhi StpA protein
            that is related to both Yersinia YopE cytotoxin and YopH tyrosine
            phosphatase
  JOURNAL   Res. Microbiol. 148 (1), 21-26 (1997)
  MEDLINE   98068312
REFERENCE   2  (bases 1 to 3964)
  AUTHORS   Popoff,M.Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-OCT-1995) M.Y. Popoff, Institut Pasteur, Unite des
            Enterobacteries, 28 rue du Docteur Roux, 75724 Paris Cedex 15,
            FRANCE
  REMARK    Revised by submittor 09-APR-96
FEATURES             Location/Qualifiers
     source          1. .3964
                     /organism="Salmonella typhi"
                     /strain="Ty2"
                     /db_xref="taxon:601"
     gene            1. .1353
                     /gene="sipA"
     CDS             <1. .1353
                     /gene="sipA"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA63302.1"
                     /db_xref="PID:e208981"
                     /db_xref="PID:g1514651"
                     /db_xref="GI:1514651"
                     /db_xref="SPTREMBL:P74849"
                     /translation="SVGREARHLEMQTLIPLLLRNVFAQIPADKLPDPKIPEPAAGPV
                     PDGGKKAEPTGINININIDSSNHSVDNSKHINNSRSHVDNSQRHIDNSNHDNSRKTID
                     NSRTFIDNSQRHGESHHSTNSSNVSHSHSRVDSTTHQTETAHSASTGTIDHGIAGKID
                     VTAHATAEAVTNSSSESKDGKVVTSEKGTTGETTSFDEVDGVTSKSIIGKPLQATVHG
                     VDDNKQQSQTAEIVNVKPLASQLAGVENVKIDTLQSDSTVITGNKAGTTDNDNSQTDK
                     TGPFSGLKFKQNSFLSTVPSVTNMHSIHFNAREAFLGVIRKALEPDASTPFPVRRAFD
                     GLRGEILPNDTIKSAALKAQCSDIDKHPELKAKMETLKEVITHHPQKEKLAEIALQFA
                     REAGLTRQKGETDYVLSNVLDGLIGDGSWRAGPAYESYLNKPGVDRVITTVDGLHMQR
                     "
     gene            1372. .1620
                     /gene="sipF"
     CDS             1372. .1620
                     /gene="sipF"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA63303.1"
                     /db_xref="PID:e208982"
                     /db_xref="PID:g1514652"
                     /db_xref="GI:1514652"
                     /db_xref="SPTREMBL:P74850"
                     /translation="MNMDIEARVKKVITSCIAVDVDSINGQTHLVEDLYPDSLDLIDI
                     VFGLSEEFDISCNENDLPDMTTFADICRVVKKSLESRV"
     gene            2330. .3961
                     /gene="stpA"
     CDS             2330. .3961
                     /gene="stpA"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA63304.1"
                     /db_xref="PID:e209029"
                     /db_xref="PID:g1514653"
                     /db_xref="GI:1514653"
                     /db_xref="SPTREMBL:P74851"
                     /translation="MLRYDERKLNNLTLSSFSKSGVSSDIRLYIAKENTDKAYVAPEK
                     FSSKVLTWLGKMPLFKNTEVVQKHTENIRVQNQKILQTFLQALTEKYGEKAVNNALYM
                     SSINMNKPLTQRLVVQITECVKGADGGFINLIKNKDNVGVMNAALVIKGGDIKVTEQN
                     NDVGAESKQPLLDIALKGLKRTIPQLEQMDGNSLRENFQEMASGNGPLRSLMTNLQSL
                     NKIPEAKQLNDYVTTLKNIQIGADRFSQWGTCGGEVERWIDKASTHELTQAVKKIHVI
                     AKELKNVTAELEKIKAGASMPQTMSGPTLGLARFAVSSIPINQQTQVKLSDGMPVPVN
                     TLTFDGKPVALAGSCPKNTPDALEAHMKMLLEKECSCLVVLTSEDQMQAKQLPAYFRG
                     SYTFGEVHTNSQKVSSASQGGAIDQYNMQLSCGEKRYTIPVLHVKNWPDHQPLPSTDQ
                     LEYLADRVKNSNQNGAPGRSSSDKHLPMIHCLGGVGRTGTMAAALVLKDNPHSNLEQV
                     RADFRNSRNNRMLEDASQFVQLKAMQAQLLMTTAS"
BASE COUNT     1264 a    759 c    966 g    975 t
ORIGIN
```

```
alignment_scores:
            Quality:  941.50              Length:    240
              Ratio:  4.222                Gaps:      3
 Percent Similarity:  92.917   Percent Identity: 90.417

alignment_block:
 US-09-068-804-12 x STSIPSTP   ..

 Align seg 1/1  to: STSIPSTP  from: 1  to: 3964

       2 ProGlyThrIleCysGlyGlnGlnHisSer.IleAsnGlnGlnThrGlnV 18
         ||||||||||:::|||||||||||||||||| |||||||||||||||||||
    3247 CCTGGCACGGTTTGCGGTCAGCAGCATTCCCATTAATCAGCAAACCCAGG 3296

      18 alLysLeuSerAspGlyMetProValProValIleArgLeuThrPheAsp 34
         |||||||||||||||||||||||||||||||      ||||||||||||
    3297 TGAAACTGAGCGATGGAATGCCTGTGCCAGTGAATACGTTAACTTTTGAC 3346

      35 GlyLysProValAlaLeuAlaGlyIleArgThrGlnLysIleArgProAs 51
         ||||||||||||||||||||||||          :::|||      |||||
    3347 GGTAAGCCTGTGGCATTAGCCGGT...TCGTGCCCAAAAAATACGCCGGA 3393

      51 pArgLeuGluAlaHisMetLysMetLeuLeuGluLysGluCysSerCysL 68
         |   ||||||||||||||||||||||||||||||||||||||||||||||||
    3394 TGCGCTGGAAGCGCATATGAAAATGCTTCTTGAAAAAGAATGTTCGTGCC 3443

      68 euValValLeuThrSerGluArgSerAspAlaGlyLysThrIleThrThr 84
         ||||||||||||||||||||||          :::|||   :::::::::
    3444 TGGTGGTGTTAACGTCGGAAGATCAGATGCAGGCAAAACAATTACCAGCC 3493

      85 TyrPheArgGlySerTyrThrPheGlyGluValHisThrAsnSerGlnLy 101
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    3494 TATTTCAGAGGGAGCTACACCTTTGGCGAGGTGCATACCAACAGCCAAAA 3543

     101 sValSerSerAlaSerGlnGlyGluAlaIleAspGlnTyrAsnMetGlnL 118
         |||||||||||||||||||||    ||||||||||||||||||||||||
    3544 AGTGAGCTCAGCCAGTCAGGGAGGAGCGATAGACCAATACAATATGCAAC 3593

     118 euSerCysGlyGluLysArgTyrThrIleProValLeuHisValLysAsn 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    3594 TGTCCTGCGGGGAAAAGCGGTATACAATCCCGGTATTGCATGTGAAAAAT 3643

     135 TrpProAspHisGlnProLeuProSerThrAspGlnLeuGluTyrLeuAl 151
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    3644 TGGCCAGATCACCAGCCGTTACCGTCTACTGATCAGTTAGAATACCTGGC 3693

     151 aAspArgValLysAsnSerAsnGlnAsnGlyAlaProGlyArgSerSerS 168
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    3694 GGATAGGGTGAAAAATAGTAACCAAAATGGCGCGCCGGGGCGTAGTTCAT 3743

     168 erAspLysHisLeuProMetIleHisCysLeuGlyGlyValGlyArgThr 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    3744 CAGATAAGCATTTACCGATGATTCATTGTCTGGGCGGAGTGGGAAGAACC 3793

     185 GlyThrMetAlaAlaAlaLeuValLeuLysAspAsnProHisSerAsnLe 201
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    3794 GGAACGATGGCGGCGGCTCTTGTACTTAAGGATAATCCTCATAGTAATCT 3843

     201 uGluGlnValArgAlaAspPheArg.IleHisGlyThrIleGluCysTrp 217
         |||||||||||||||||||||||| ||||||||||||||||||||||||
    3844 GGAGCAGGTACGTGCAGATTTCCGGAATTCACGGAACAATAGAATGCTGG 3893

     218 LysThrProAlaGlnPheValGlnLeuLysAlaMetGlnAlaGlnLeuLe 234
         |||||||||   ||||||||||||||||||||||||||||||||||||||
    3894 AAGACGCCTCA.CAGTTTGTACAACTAAAAGCAATGCAAGCCCAGTTGCT 3942

     234 uMetThrThrAlaSer 239
         |||||||||||||||
    3943 TATGACGACGGCAAGC 3958

seq_name: gb_ba1:YEPYOP51A

seq_documentation_block:
LOCUS       YEPYOP51A    2000 bp    DNA             BCT       19-JUL-1994
DEFINITION  Yersinia enterocolitica YopH protein gene, complete cds.
ACCESSION   M30457
NID         g155530
VERSION     M30457.1  GI:155530
KEYWORDS    YopH protein.
SOURCE      Yersinia enterocolitica (strain W22703) DNA.
  ORGANISM  Yersinia enterocolitica
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Yersinia.
REFERENCE   1  (bases 1 to 2000)
  AUTHORS   Michiels,T. and Cornelis,G.
  TITLE     Nucleotide sequence and transcription analysis of yop51 from
            Yersinia enterocolitica W22703
  JOURNAL   Microb. Pathog. 5 (6), 449-459 (1988)
  MEDLINE   89218619
REFERENCE   2  (bases 1 to 2000)
  AUTHORS   Michiels,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-DEC-1989) T. Michiels, Unite de Microbiologie,
            Universite Catholique de Louvain, Brussels, Belgium
FEATURES             Location/Qualifiers
     source          1. .2000
                     /organism="Yersinia enterocolitica"
                     /strain="W22703"
                     /isolate="439 80"
                     /db_xref="taxon:630"
     -10_signal      273. .278
     RBS             312. .319
                     /note="ribosome binding site (put.); putative"
     gene            325. .1731
                     /gene="yopH"
     CDS             325. .1731
                     /gene="yopH"
                     /note="formerly called yop51"
                     /codon_start=1
                     /transl_table=11
                     /product="YopH protein"
                     /protein_id="AAA19860.1"
                     /db_xref="PID:g155531"
                     /db_xref="GI:155531"
                     /translation="MNLSLSDLHRQVSRLVQQESGDCTGKLRGNVAANKETTFQGLTI
                     ASGARESEKVFAQTVLSHVANIVLTQEDTAKLLQSTVKHNLNNYELRSVGNGNSVLVS
                     LRSDQMTLQDAKVLLEAALRQESGARGHVSSHSHSVLHAPGTPVREGLRSHLDPRTPP
                     LPPRERPHTSGHHGAGEARATAPSTVSPYGPEARAELSSRLTTLRNTLAPATNDPRYL
                     QACGGEKLNRFRDIQCCRQTAVRADLNANYIQVGNTRTIACQYPLQSQLESHFRMLAE
                     NRTPVLAVLASSSEIANQRFGMPDYFRQSGTYGSITVESKMTQQVGLGDGIMADMYTL
                     TIREAGQKTISVPVVHVGNWPDQTAVSSEVTKALASLVDQTAETKRNMYESKGSSAVA
                     DDSKLRPVIHCRAGVGRTAQLIGAMCMNDSRNSQLSVEDMVSQMRVQRNGIMVQKDEQ
                     LDVLIKLAEGQGRPLLNS"
     misc_signal     734. .774
                     /gene="yopH"
                     /note="internal transcription termination signal (put.);
                     putative"
     misc_signal     1804. .1840
                     /note="transcription termination signal (put.); putative"
BASE COUNT      656 a    396 c    422 g    526 t
ORIGIN


alignment_scores:
            Quality:  202.50              Length:    200
              Ratio:  1.489                Gaps:      7
 Percent Similarity:  68.000   Percent Identity: 29.500

alignment_block:
 US-09-068-804-12 x YEPYOP51A   ..

 Align seg 1/1  to: YEPYOP51A  from: 1  to: 2000

      29 IleArgLeuThrPheAspGlyLysProValAlaLeuAlaGlyIleArgTh 45
```

```
     :::|||      ::::::::::::   :::   ::::::::::::|||||
  1042 GTACGCGCCGATCTTAATGCCAATTACATCCAGGTCGGTAACACTCGTAC 1091

    45 r.........GlnLysIleArgProAspArgLeuGluAlaHisMetLysM 59
       |         |||          :::::::::||||||:::|||   :::|
  1092 CATAGCGTGCCAGTATCCGCTACAATCTCAACTTGAAAGCCATTTCCGTA 1141

    59 etLeuLeuGluLysGluCysSerCysLeuValValLeuThrSerGluArg 75
       |||||   |||:::     :::   |||:::|||||||:::|||
  1142 TGCTGGCAGAAAACCGAACGCCAGTGTTGGCTGTTTTAGCGTCCAGTTCT 1191

    76 SerAspAlaGlyLys......ThrIleThrThrTyrPheArgGlySerTy 90
          |||::::::         ::::::   |||||||||   |||
  1192 GAGATAGCCAATCAAAGATTCGGTATGCCAGATTATTTCCGCCAGAGTGG 1241

    90 rThrPheGlyGluValHisThrAsnSerGlnLysValSerSerAlaSerG 107
        |||:::|||   :::   ::::::|||:::::::::      ::::::
  1242 TACCTATGGCAGTATCACTGTAGAGTCTAAAATGACTCAGCAAGTTGGTC 1291

   107 lnGlyGluAlaIle......AspGlnTyrAsnMetGlnLeu...SerCys 120
         |||::::::|||      |||   |||::::::   :::
  1292 TCGGTGACGGGATTATGGCAGATATGTATACTTTAACGATTCGTGAAGCG 1341

   121 GlyGluLysArgTyrThrIleProValLeuHisValLysAsnTrpProAs 137
       |||:::|||      ::::::||||||:::|||||   ||||||||||
  1342 GGTCAAAAAACAATTTCTGTTCCTGTGGTTCATGTTGGCAATTGGCCCGA 1391

   137 pHisGlnProLeuProSer......ThrAspGlnLeuGluTyrLeuAlaA 152
       |:::   :::::::::|||      |||      |||:::   |||:::|
  1392 TCAGACCGCAGTCAGCTCTGAAGTTACCAAGGCACTCGCTTCACTGGTAG 1441

   152 spArgValLysAsnSerAsnGlnAsn..............GlyAlaPro 163
       ||:::::::   ::::::::::::|||              |||::::::
  1442 ATCAAACAGCAGAAACAAAACGCAATATGTATGAAAGCAAAGGAAGTTCA 1491

   164 GlyArgSerSerSerAspLysHisLeuProMetIleHisCysLeuGlyGl 180
       :::   ::::::::::::|||      |||:::||||||||||   :::||
  1492 GCGGTAGCAGATGACTCCAAATTACGGCCGGTAATACATTGCCGTGCGGG 1541

   180 yValGlyArgThrGlyThrMetAlaAlaAlaLeuValLeuLysAspAsnP 197
       |||||||||||||:::   :::   :::|||:::   ::::::|||:::
  1542 TGTTGGCCGTACTGCGCAACTGATTGGCGCAATGTGCATGAATGATAGTC 1591

   197 roHisSerAsnLeu......GluGlnValArgAlaAspPheArgIleHis 211
         :::|||:::|||      |||::::::   ::::::   |||::::::
  1592 GTAATAGTCAGTTAAGCGTAGAAGATATGGTCAGCCAAATGCGAGTACAA 1641

seq_name: gb_ba1:YPYOPH

seq_documentation_block:
LOCUS       YPYOPH       2924 bp    DNA              BCT       17-FEB-1995
DEFINITION  Yersinia pseudotuberculosis plasmid encoded yopH gene for protein
            Yop2b.
ACCESSION   Y00551
NID         g677867
VERSION     Y00551.1  GI:677867
KEYWORDS    membrane protein; virulence protein; Yop2b protein; yopH gene.
SOURCE      Yersinia pseudotuberculosis.
  ORGANISM  Yersinia pseudotuberculosis
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Yersinia.
REFERENCE   1  (bases 1 to 2924)
  AUTHORS   Bolin,I. and Wolf-Watz,H.
  TITLE     The plasmid-encoded Yop2b protein of Yersinia pseudotuberculosis is
            a virulence determinant regulated by calcium and temperature at the
            level of transcription
  JOURNAL   Mol. Microbiol. 2 (2), 237-245 (1988)
  MEDLINE   88246051
COMMENT     On Feb 22, 1995 this sequence version replaced gi:48642.
FEATURES             Location/Qualifiers
     source          1. .2924
                     /organism="Yersinia pseudotuberculosis"
                     /plasmid=""
                     /strain="YPIII (pIB1)"
                     /db_xref="taxon:633"
     promoter        748. .775
                     /note="-35 and -10 region"
     precursor_RNA   783. .>783
     CDS             806. .2212
                     /codon_start=1
                     /transl_table=11
                     /product="Yop2b protein"
                     /protein_id="CAA68629.1"
                     /db_xref="PID:g677868"
                     /db_xref="GI:677868"
                     /db_xref="SWISS-PROT:P08538"
                     /translation="MNLSLSDLHRQVSRLVQQESGDCTGKLRGNVAANKETTFQGLTI
                     ASGARESEKVFAQTVLSHVANVVLTQEDTAKLLQSTVKHNLNNYDLRSVGNGNSVLVS
                     LRSDQMTLQDAKVLLEAALRQESGARGHVSSHSHSALHAPGTPVREGLRSHLDPRTPP
                     LPPRERPHTSGHHGAGEARATAPSTVSPYGPEARAELSSRLTTLRNTLAPRTNDPRYL
                     QACGGEKLNRFRDIQCCRQTAVRADLNANYIQVGNTRTIACQYPLQSQLESHFRMLAE
                     NRTPVLAVLASSSEIANQRFGMPDYFRQSGTYGSITVESKMTQQVGLGDGIMADMYTL
                     TIREAGQKTISVPVVHVGNWPDQTAVSSEVTKALASLVDQTAETKRNMYESKGSSAVG
                     DDSKLRPVIHCRAGVGRTAQLIGAMCMNDSRNSQLSVEDMVSQMRVQRNGIMVQKDEQ
                     LDVLIKLAEGQGRPLLNS"
BASE COUNT      842 a    689 c    648 g    745 t
ORIGIN


alignment_scores:
            Quality:  202.50            Length:    200
              Ratio:  1.489               Gaps:      7
 Percent Similarity:  68.000   Percent Identity: 29.500

alignment_block:
 US-09-068-804-12 x YPYOPH   ..

 Align seg 1/1  to: YPYOPH  from: 1  to: 2924

     29 IleArgLeuThrPheAspGlyLysProValAlaLeuAlaGlyIleArgTh 45
        :::|||      ::::::::::::   :::   ::::::::::::|||||
   1523 GTACGCGCCGATCTTAATGCCAATTACATCCAGGTCGGTAACACTCGTAC 1572

     45 r.........GlnLysIleArgProAspArgLeuGluAlaHisMetLysM 59
        |         |||          :::::::::||||||:::|||   :::|
   1573 CATAGCGTGCCAGTATCCGCTACAATCTCAACTTGAAAGCCATTTCCGTA 1622

     59 etLeuLeuGluLysGluCysSerCysLeuValValLeuThrSerGluArg 75
        |||||   |||:::     :::   |||:::|||||||:::|||
   1623 TGCTGGCAGAAAACCGAACGCCAGTGTTGGCTGTTTTAGCGTCCAGTTCT 1672

     76 SerAspAlaGlyLys......ThrIleThrThrTyrPheArgGlySerTy 90
           |||::::::         ::::::   |||||||||   |||
   1673 GAGATAGCCAATCAAAGATTCGGTATGCCAGATTATTTCCGCCAGAGTGG 1722

     90 rThrPheGlyGluValHisThrAsnSerGlnLysValSerSerAlaSerG 107
         |||:::|||   :::   ::::::|||:::::::::      ::::::
   1723 TACCTATGGCAGTATCACTGTAGAGTCTAAAATGACTCAGCAAGTTGGTC 1772

    107 lnGlyGluAlaIle......AspGlnTyrAsnMetGlnLeu...SerCys 120
          |||::::::|||      |||   |||::::::   :::
   1773 TCGGTGACGGGATTATGGCAGATATGTATACTTTAACGATTCGTGAAGCG 1822

    121 GlyGluLysArgTyrThrIleProValLeuHisValLysAsnTrpProAs 137
        |||:::|||      ::::::||||||:::|||||   ||||||||||
   1823 GGTCAAAAAACAATCTCTGTTCCTGTGGTTCATGTTGGCAATTGGCCCGA 1872

    137 pHisGlnProLeuProSer......ThrAspGlnLeuGluTyrLeuAlaA 152
        |:::   :::::::::|||      |||      |||:::   |||:::|
   1873 TCAGACCGCAGTCAGCTCTGAAGTTACCAAGGCACTCGCTTCACTGGTAG 1922

    152 spArgValLysAsnSerAsnGlnAsn..............GlyAlaPro 163
        ||:::::::   ::::::::::::|||              |||::::::
   1923 ATCAAACAGCAGAAACAAAACGCAATATGTATGAAAGCAAAGGAAGTTCA 1972

    164 GlyArgSerSerSerAspLysHisLeuProMetIleHisCysLeuGlyGl 180
```

```
      :::   :::::::::::::|||      |||:::||||||||||   :::||
   1973 GCGGTAGGAGATGACTCCAAATTACGGCCGGTAATACATTGCCGTGCGGG 2022

    180 yValGlyArgThrGlyThrMetAlaAlaAlaLeuValLeuLysAspAsnP 197
        ||||||||||||||:::   :::   :::|||:::   :::::::|||:::
   2023 TGTIGGCCGTACTGCGCAACTGATTGGCGCAATGTGCATGAATGATAGTC 2072

    197 roHisSerAsnLeu......GluGlnValArgAlaAspPheArgIleHis 211
          :::|||:::|||      |||::::::   ::::::   |||::::::
   2073 GTAATAGTCAGTTAAGCGTAGAAGATATGGTCAGCCAAATGCGAGTACAA 2122
```

seq_name: gb_ba2:AF102990

seq_documentation_block:

LOCUS AF102990 69673 bp DNA circular BCT 18-MAY-1999
DEFINITION Yersinia enterocolitica plasmid pYVe227, complete sequence.
ACCESSION AF102990 AF054978 AF054979 AF054980 AF054981 AF080156 Z69926 L06216 AF080155 AF022645 AF050104 AF054977 AF033863 U02499 U08019 U21297 M22781 U08222 Z18539 U94827
NID g4324323
VERSION AF102990.1 GI:4324323
KEYWORDS .
SOURCE Yersinia enterocolitica.
ORGANISM Yersinia enterocolitica
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Yersinia.
REFERENCE 1 (bases 32056 to 32871)
AUTHORS Cornelis,G., Sluiters,C., de Rouvroit,C.L. and Michiels,T.
TITLE Homology between virF, the transcriptional activator of the Yersinia virulence regulon, and AraC, the Escherichia coli arabinose operon regulator
JOURNAL J. Bacteriol. 171 (1), 254-262 (1989)
MEDLINE 89123026
REFERENCE 2 (bases 15074 to 15994)
AUTHORS Hakansson,S., Bergman,T., Vanooteghem,J.C., Cornelis,G. and Wolf-Watz,H.
TITLE YopB and YopD constitute a novel class of Yersinia Yop proteins
JOURNAL Infect. Immun. 61 (1), 71-80 (1993)
MEDLINE 93114907
REFERENCE 3 (bases 45325 to 45717)
AUTHORS Wattiau,P. and Cornelis,G.R.
TITLE SycE, a chaperone-like protein of Yersinia enterocolitica involved in Ohe secretion of YopE
JOURNAL Mol. Microbiol. 8 (1), 123-131 (1993)
MEDLINE 93268087
REFERENCE 4 (bases 23950 to 25269)
AUTHORS Woestyn,S., Allaoui,A., Wattiau,P. and Cornelis,G.R.
TITLE YscN, the putative energizer of the Yersinia Yop secretion machinery
JOURNAL J. Bacteriol. 176 (6), 1561-1569 (1994)
MEDLINE 94179088
REFERENCE 5 (bases 29897 to 30961)
AUTHORS Allaoui,A., Woestyn,S., Sluiters,C. and Cornelis,G.R.
TITLE YscU, a Yersinia enterocolitica inner membrane protein involved in Yop secretion
JOURNAL J. Bacteriol. 176 (15), 4534-4542 (1994)
MEDLINE 94321323
REFERENCE 6 (bases 43292 to 43717)
AUTHORS Wattiau,P., Bernier,B., Deslee,P., Michiels,T. and Cornelis,G.R.
TITLE Individual chaperones required for Yop secretion by Yersinia
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 91 (22), 10493-10497 (1994)
MEDLINE 95024141
REFERENCE 7 (bases 31537 to 31932)
AUTHORS Allaoui,A., Scheen,R., Lambert de Rouvroit,C. and Cornelis,G.R.
TITLE VirG, a Yersinia enterocolitica lipoprotein involved in Ca2+ dependency, is related to exsB of Pseudomonas aeruginosa
JOURNAL J. Bacteriol. 177 (15), 4230-4237 (1995)
MEDLINE 95362644
REFERENCE 8 (bases 53135 to 53497)
AUTHORS Stainier,I., Iriarte,M. and Cornelis,G.R.
TITLE YscM1 and YscM2, two Yersinia enterocolitica proteins causing downregulation of yop transcription
JOURNAL Mol. Microbiol. 26 (4), 833-843 (1997)
MEDLINE 98087284
REFERENCE 9 (bases 18681 to 18968)
AUTHORS Boyd,A.P., Sory,M.P., Iriarte,M. and Cornelis,G.R.
TITLE Heparin interferes with translocation of Yop proteins into HeLa cells and binds to LcrG, a regulatory component of the Yersinia Yop apparatus
JOURNAL Mol. Microbiol. 27 (2), 425-436 (1998)
MEDLINE 98143428
REFERENCE 10 (bases 22612 to 22890)
AUTHORS Iriarte,M., Sory,M.P., Boland,A., Boyd,A.P., Mills,S.D., Lambermont,I. and Cornelis,G.R.
TITLE TyeA, a protein involved in control of Yop release and in translocation of Yersinia Yop effectors
JOURNAL EMBO J. 17 (7), 1907-1918 (1998)
MEDLINE 98190073
REFERENCE 11 (bases 1 to 69673)
AUTHORS Iriarte,M., Lambermont,I., Kerbourch,C. and Cornelis,G.R.
TITLE Detailed genetic map of the pYVe227 plasmid of Yersinia enterocolitica serotype O:9
JOURNAL Unpublished
REFERENCE 12 (bases 1 to 69673)
AUTHORS Iriarte,M., Lambermont,I., Kerbourch,C. and Cornelis,G.R.
TITLE Direct Submission
JOURNAL Submitted (30-OCT-1998) Microbial Pathogenesis Unit, Christian de Duve Institute of Cellular Pathology and Faculte de Medecine. Universite Catholique de Louvain, Av. Hippocrate 74.49, Brussels 1200, Belgium
COMMENT On May 18, 1999 this sequence version replaced gi:3411136 gi:3411135 gi:3411139 gi:3411137 gi:3603360 gi:1524380 gi:155536 gi:3603357 gi:2738986 gi:2944089 gi:3411138 gi:3108216 gi:437201 gi:497215 gi:885971 gi:155526 gi:483322 gi:48593 gi:2735421.

FEATURES Location/Qualifiers

```
     source          1. .69673
                     /organism="Yersinia enterocolitica"
                     /strain="W22703"
                     /serotype="O:9"
                     /db_xref="taxon:630"
                     /plasmid="pYVe227"
     rep_origin      53. .203
     CDS             complement(889. .1434)
                     /note="ORF181"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAD16801.1"
                     /db_xref="PID:g4324324"
                     /db_xref="GI:4324324"
                     /translation="MRVVIYQGQYGIFKKNGLDLSMPAIQTCLGLYALSDQHDYLLCA
                     HFDSAFGLQQNLQDIKRFIELNGIRMSSLRATVFGGDGKQSYLRCSAPSSHIGNEIVN
                     FMKSHGALAQYSNQYYSGVFPETFNFHYQKGCGITKGQNPRDFLGGSPQAMNLAKQRI
                     KLRPSEYSLTHAKMLDISRLY"
     CDS             1616. .2083
                     /note="ORF155"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAD16802.1"
                     /db_xref="PID:g4324325"
                     /db_xref="GI:4324325"
                     /translation="MINTTFTELLQKIASHFGLDKLSQDEYGLCELILNDRVVIMLRA
                     DEILNRLTLLGPILGFSGPEARSTASQLFFCYSINALNKDGPCFAWSEELGLIAFKHL
                     SLGELNVENVSKEIANFYDWLSLVSLPAETQQELPPHTQSTQSTQSTQSVKWG"
     gene            2091. .4280
                     /gene="yopO"
     CDS             2091. .4280
                     /gene="yopO"
                     /codon_start=1
                     /transl_table=11
                     /product="protein kinase YopO"
                     /protein_id="AAD16803.1"
                     /db_xref="PID:g4324326"
                     /db_xref="GI:4324326"
                     /translation="MKIMGTMPPSISLAKAHERISQHWQNPVGELNIGGKRYRIIDNQ
```

```
                VLRLNPHSGFSLFREGVGKIFSGKMFNFSIARNLTETLHAAQKTTSQELRSDIPNALS
                NLFGAKPQTELPLGWKGKPLSGAPDLEGMRVAETDKFAEGESHISIIETKDNQRLVAK
                IERSIAEGHLFAELEAYKHIYKTAGKHPNLANVHGMAVVPYGNRKEEALLMDEVDGWR
                CSDTLRSLADSWKQGKINSEAYWGTIKFIAHRLLDVTNHLAKAGIVHNDIKPGNVVFD
                RASGEPVVIDLGLHSRSGEQPKGFTESFKAPELGVGNLGASEKSDVFLVVSTLLHGIE
                GFEKDPEIKPNQGLRFITSEPAHVMDENGYPIHRPGIAGVETAYTRFITDILGVSADS
                RPDSNEARLHEFLSDGTIDEESAKQILKDTLTGEMSPLSTDVRRITPKKLRELSDLLR
                THLSSAATKQLDMGVVLSDLDTMLVTLDKAERGGVDKDQLKSFNSLILKTYSVIEDY
                VKGREGDTKSSSAEVSPYHRSNFMLSIAEPSLQRIQKHLDQTHSFSDIGSLVRAHKHL
                ETLLEVLVTLSPQGQPVSSETYSFLNRLAEAKVTLSQQLDTLQQQQESAKAQLSILIN
                RSGSWADVARQSLQRFDSTRPVVKFGTEQYTAIHRQMMAAHAAITLQEVSEFTDDMRN
                FTADSIPLLIRLGRSSLIDEHLVEQREKLRELTTIAERLNRLEREWM"
     CDS        4485. .4760
                /note="ORF91A; unknown function"
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AAD16804.1"
                /db_xref="PID:g4324327"
                /db_xref="GI:4324327"
                /translation="MLSIRKNECPCWTLNNGLPVICSGELAINEGKFVGINNYSALYR
                IMNLGVLVRVLPVCSGFTHQRLNVLLQGTVYPEYKIHTAVNSLNKDK"
     gene       4761. .5627
                /gene="yopP"
     CDS        4761. .5627
                /gene="yopP"
                /codon_start=1
                /transl_table=11
                /product="Yop effector YopP"
                /protein_id="AAD16805.1"
                /db_xref="PID:g4324328"
                /db_xref="GI:4324328"
                /translation="MIGPISQINSFGGLSEKETRSLISNEELKNIIIQLETDIADGSW
                FHKNYSRLDIEVMPALVIQANNKYPEMNLNFVTSPQDLSIEIKNVIENGVGSSRFIIN
                MGEGGIHFSVIDYKHINGKTSLILFEPVNFNSMGPAILAISTKTAIERYQLPDCHFSM
                VEMDIQRSSSECGIFSLALAKKLYTERDSLLKIHEDNIKGILSDSENPLPHNKLDPYL
                PVTFYKHTQGKKRLNEYLNTNPQGVGTVVNKKNETIFNRFDNNKSIIDGKELSVSVHK
                KRIAEYKTLLKV"
     gene       complement(6151. .6900)
                /gene="ylpA"
     CDS        complement(6151. .6900)
                /gene="ylpA"
                /codon_start=1
                /transl_table=11
                /product="lipoprotein YlpA"
                /protein_id="AAD16806.1"
                /db_xref="PID:g4324329"
                /db_xref="GI:4324329"
                /translation="MEDDMKKNMKLIAITAVLSSVLVLSGCGAMSTAIKKRNLEVKTQ
                MSETIWLEPSSQKTVYLQIKNTSDKNMLGLAPKITKAVQDKGYTVTSSPEDAHYWIQA
                NVLKADKMDLREAEGFLSQGYQGAALGAALGAGITGYNSNSAGASLGVGLAAGLVGMV
                ADAMVEDINYTMVTDVQISEKTDTPLQTDNVAALKQGTSGYKVQTSTQTGNKHQYQTR
                VVSSANKVNLKFEEAQPVLEDQLAKSIANIL"
     gene       complement(7370. .7918)
                /gene="yopQ"
     CDS        complement(7370. .7918)
                /gene="yopQ"
                /codon_start=1
                /transl_table=11
                /product="YopQ"
                /protein_id="AAD16807.1"
                /db_xref="PID:g4324330"
                /db_xref="GI:4324330"
                /translation="MFIKDAYNMRALCTALEQSAPDTIINTSKEENNSYYCATAHLLR
                TDVCSLVNRVGIEPLKSGSILSTLEELWQAVGIIYRLYEWQHVSDIDTNFKKLPNNSD
                FGLVFSVLDCDIGYVITGKKDSKGNIELYDPKNSLLIENDDIKKYLYDENFHRFCIML
                IISKSELEELSRESCDQKCIMG"
     gene       8439. .9407

alignment_scores:
            Quality:  202.50          Length:    200
              Ratio:  1.489             Gaps:      7
 Percent Similarity:  68.000  Percent Identity: 29.500
```

```
alignment_block:
 US-09-068-804-12 x AF102990   ..

 Align seg 1/1  to: AF102990  from: 1  to: 69673

      29 IleArgLeuThrPheAspGlyLysProValAlaLeuAlaGlyIleArgTh 45
         :::|||      ::::::::::::   :::   ::::::::::::|||||
   42375 GTACGCGCCGATCTTAATGCCAATTACATCCAGGTCGGTAACACTCGTAC 42424

      45 r.........GlnLysIleArgProAspArgLeuGluAlaHisMetLysM 59
         |         |||          :::::::::|||||:::|||   :::|
   42425 CATAGCGTGCCAGTATCCGCTACAATCTCAACTTGAAAGCCATTTCCGTA 42474

      59 etLeuLeuGluLysGluCysSerCysLeuValValLeuThrSerGluArg 75
         |||||   |||:::      :::   |||:::||||||:::|||
   42475 TGCTGGCAGAAAACCGAACGCCAGTGTTGGCTGTTTTAGCGTCCAGTTCT 42524

      76 SerAspAlaGlyLys......ThrIleThrThrTyrPheArgGlySerTy 90
            |||::::::          ::::::   |||||||||   |||
   42525 GAGATAGCCAATCAAAGATTCGGTATGCCAGATTATTTCCGCCAGAGTGG 42574

      90 rThrPheGlyGluValHisThrAsnSerGlnLysValSerSerAlaSerG 107
          |||:::|||   :::   ::::::|||:::::::::      ::::::
   42575 TACCTATGGCAGTATCACTGTAGAGTCTAAAATGACTCAGCAAGTTGGTC 42624

     107 lnGlyGluAlaIle......AspGlnTyrAsnMetGlnLeu...SerCys 120
           |||:::::|||      |||   |||::::::   :::
   42625 TCGGTGACGGGATTATGGCAGATATGTATACTTTAACGATTCGTGAAGCG 42674

     121 GlyGluLysArgTyrThrIleProValLeuHisValLysAsnTrpProAs 137
         |||:::|||      ::::::|||||::|||||   ||||||||||
   42675 GGTCAAAAAACAATTTCTGTTCCTGTGGTTCATGTTGGCAATTGGCCCGA 42724

     137 pHisGlnProLeuProSer......ThrAspGlnLeuGluTyrLeuAlaA 152
         |:::   :::::::::|||      |||      |||:::   |||:::|
   42725 TCAGACCGCAGTCAGCTCTGAAGTTACCAAGGCACTCGCTTCACTGGTAG 42774

     152 spArgValLysAsnSerAsnGlnAsn.............GlyAlaPro 163
         ||::::::   ::::::::::::|||              |||::::::
   42775 ATCAAACAGCAGAAACAAAACGCAATATGTATGAAAGCAAAGGAAGTTCA 42824

     164 GlyArgSerSerSerAspLysHisLeuProMetIleHisCysLeuGlyGl 180
         :::   :::::::::::::|||      |||:::|||||||||   ::::|
   42825 GCGGTAGCAGATGACTCCAAATTACGGCCGGTAATACATTGCCGTGCGGG 42874

     180 yValGlyArgThrGlyThrMetAlaAlaAlaLeuValLeuLysAspAsnP 197
         |||||||||||||:::   :::   ::::|||:::   :::::::|||:::
   42875 TGTTGGCCGTACTGCGCAACTGATTGGCGCAATGTGCATGAATGATAGTC 42924

     197 roHisSerAsnLeu......GluGlnValArgAlaAspPheArgIleHis 211
           :::|||:::|||      |||::::::   ::::::   |||::::::
   42925 GTAATAGTCAGTTAAGCGTAGAAGATATGGTCAGCCAAATGCGAGTACAA 42974

 seq_name: gb_ba2:AF053946

 seq_documentation_block:
 LOCUS       AF053946    70504 bp    DNA   circular  BCT       06-OCT-1998
 DEFINITION  Yersinia pestis plasmid pCD1, complete plasmid sequence.
 ACCESSION   AF053946
 NID         g2996222
 VERSION     AF053946.1  GI:2996222
 KEYWORDS    .
 SOURCE      Yersinia pestis.
   ORGANISM  Yersinia pestis
             Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
             Yersinia.
 REFERENCE   1  (bases 1 to 70504)
   AUTHORS   Hu,P., Elliott,J., McCready,P., Skowronski,E., Garnes,J.,
             Kobayashi,A., Brubaker,R.R. and Garcia,E.
   TITLE     Structural organization of virulence-associated plasmids of
             Yersinia pestis
   JOURNAL   J. Bacteriol. 180 (19), 5192-5202 (1998)
```

```
  MEDLINE   98422474
REFERENCE   2  (bases 1 to 70504)
  AUTHORS   Hu,P., Elliott,J., McCready,P., Skowronski,E., Garnes,J.,
            Kobayashi,A., Carrano,A.V., Brubaker,R. and Garcia,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAR-1998) Biology and Biotechnology, Lawrence
            Livermore National Lab, L452, 7000 East Ave., Livermore, CA 94550,
            USA
FEATURES             Location/Qualifiers
     source          1. .70504
                     /organism="Yersinia pestis"
                     /plasmid="pCD1"
                     /strain="KIM"
                     /db_xref="taxon:632"
     CDS             87. .1109
                     /codon_start=1
                     /transl_table=11
                     /product="transposase"
                     /protein_id="AAC62546.1"
                     /db_xref="PID:g2996223"
                     /db_xref="GI:2996223"
                     /translation="MVTFETVMEIKILHKQGMSSRAIARELGISRNTVKRYLQAKSEP
                     PKYTPRPAVASLLDEYRDYIRQRIADAHPYKIPATVIAREIRDQGYRGGMTILRAFIR
                     SLSVPQEQEPAVRFETEPGRQMQVDWGTMRNGRSPLHVFVAVLGYSRMLYIEFTDNMR
                     YDTLETCHRNAFRFFGGVPREVLYDNMKTVVLQRDAYQTGQHRFHPSLWQFGKEMGFS
                     PRLCRPFRAQTKGKVERMVQYTRNSFYIPLMTRLRPMGITVDVETANRHGLRWLHDVA
                     NQRKHETIQARPCDRWLEEQQSMLALPPEKKEYDVHLDENLVNFDKHPLHHPLSIYDS
                     FCRGVA"
     CDS             1109. .1888
                     /codon_start=1
                     /transl_table=11
                     /product="transposase"
                     /protein_id="AAC62557.1"
                     /db_xref="PID:g2996234"
                     /db_xref="GI:2996234"
                     /translation="MMELQHQRLMALAGQLQLESLISAAPALSQQAVDQEWSYMDFLE
                     HLLHEEKLARHQRKQAMYTRMAAFPAVKTFEEYDFTFATGAPQKQLQSLRSLSFIERN
                     ENIVLLGPSGVGKTHLAIAMGYEAVRAGIKVRFTTAADLLLQLSTAQRQGRYKTTLQR
                     GVMAPRLLIIDEIGYLPFSQEEAKLFFQVIAKRYEKSAMILTSNLPFGQWDQTFAGDA
                     ALTSAMLDRILHHSHVVQIKGESYRLRQKRKAGVIAEANPE"
     CDS             complement(1939. .2343)
                     /codon_start=1
                     /transl_table=11
                     /product="transposase"
                     /protein_id="AAC62568.1"
                     /db_xref="PID:g2996245"
                     /db_xref="GI:2996245"
                     /translation="MTEYQASERRGCRIMGISRSLLHYCPNTARDIPVVEVLQKLAHQ
                     YPAYGFGLMFNKLRQSGLPWNVKRVYRVYRLLKLNFRRKGKKRLPNRHPQPLAIPLKM
                     NHCWSVDFMSDALTDGRRFRLFNVVEIFVNDG"
     gene            complement(2379. .2645)
                     /gene="lcrS"
     CDS             complement(2379. .2645)
                     /gene="lcrS"
                     /codon_start=1
                     /transl_table=11
                     /product="low calcium response locus protein S homolog"
                     /protein_id="AAC62579.1"
                     /db_xref="PID:g2996256"
                     /db_xref="GI:2996256"
                     /translation="MKKARFTETQILRVLKEVEGGRHVKDVCRENGGSEASYYNWKSK
                     YGGMESSDIKRMKEREEENRRLKQMYASLSLDHEILKDVVAKKL"
     gene            complement(3193. .3540)
                     /gene="yscM"
     CDS             complement(3193. .3540)
                     /gene="yscM"
                     /codon_start=1
                     /transl_table=11
                     /product="Yop proteins translocation protein M homolog"
                     /protein_id="AAC62586.1"
                     /db_xref="PID:g2996263"
                     /db_xref="GI:2996263"
                     /translation="MKINTLQSLINQQITQVGHGGQAGRLTETNPLTENSHQISTAEK
                     AFANEVLEHVKNTALSRHDIACLLPRVSNLELKQGKAGEVIVTGLRTEQLSLSDAKLL
                     LEAAMRQDTAADG"
     gene            complement(3765. .4430)
                     /gene="yscL"
     CDS             complement(3765. .4430)
                     /gene="yscL"
                     /codon_start=1
                     /transl_table=11
                     /product="Yop proteins translocation protein L  homolog"
                     /protein_id="AAC62596.1"
                     /db_xref="PID:g2996273"
                     /db_xref="GI:2996273"
                     /translation="MSQTCQTGYAYMQPFVQIIPSNLSLACGLRILRAEDYQSSLTTE
                     ELISAAKQDAEKILADAQEVYEQQKQLGWQAGMDEARTLQATLIHETQLQCQQFYRHV
                     EQQMSEVVLLAVRKILNDYDQVAMTLQVVREALALVSNQKQVVVRVNPDQAGAIREQI
                     AKVHKDFPEISYLEVTADARLDQGGCILETEVGIIDASIDGQIEALSRAISTTLGQMK
                     VTE"
     gene            complement(4376. .5005)
                     /gene="yscK"
     CDS             complement(4376. .5005)
                     /gene="yscK"
                     /codon_start=1
                     /transl_table=11
                     /product="Yop proteins translocation protein K homolog"
                     /protein_id="AAC62605.1"
                     /db_xref="PID:g2996282"
                     /db_xref="GI:2996282"
                     /translation="MMENYITSFQLRFCPAAYLHLEQLPSLWRSILPYLPQWRDSAHL
                     NAALLDEFSLDTDYEEPHGLGALPLQPQSQLELLLCRLGLVLHGEAIRRCVLASPLQQ
                     LLTLVNQETLRQIIVQHELLIGPWPTHWQRPLPTEIESRTMIQSGLAFWLAAMEPQPQ
                     AWCKRLSLRLPLATPSEPWLVAESQRPLAQTLCHKLVKQVTPTCSHLFK"
     gene            complement(5005. .5739)
                     /gene="yscJ"
     CDS             complement(5005. .5739)
                     /gene="yscJ"
                     /codon_start=1
                     /transl_table=11
                     /product="Yop proteins translocation lipoprotein J"
                     /protein_id="AAC62607.1"
                     /db_xref="PID:g2996284"
                     /db_xref="GI:2996284"
                     /translation="MKVKTSLSTLILILFLTGCKVDLYTGISQKEGNEMLALLRQEGL
                     SADKEPDKDGKIKLLVEESDVAQAIDILKRKGYPHESFSTLQDVFPKDGLISSPIEEL
                     ARLNYAKAQEISRTLSEIDGVLVARVHVVLPEEQNNKGKKGVAASASVFIKHAADIQF
                     DTYIPQIKQLVNNSIEGLAYDRISVILVPSVDVRQSSHLPRNTSILSIQVSEESKGHL
                     IGLLSLLILLLPVTNLAQYFWLQRKK"
     gene            complement(5746. .6093)
                     /gene="yscI"
     CDS             complement(5746. .6093)
                     /gene="yscI"
                     /codon_start=1
                     /transl_table=11
                     /product="Yop proteins translocation protein I homolog"
                     /protein_id="AAC62608.1"
                     /db_xref="PID:g2996285"
                     /db_xref="GI:2996285"
                     /translation="MPNIEIAQADEVIITTLEELGPAEPTTDQIMRFDAAMSEDTQGL
                     GHSLLKEVSDIQKSFKTVKSDLHTKLAVSVDNPNDLMLMQWSLIRITIQEELIAKIAG
                     RMSQNVETLSKGG"
     gene            complement(6094. .6591)
                     /gene="yscH"
     CDS             complement(6094. .6591)
                     /gene="yscH"
                     /codon_start=1
                     /transl_table=11
                     /product="Yop proteins translocation protein H homolog"
                     /protein_id="AAC62547.1"
                     /db_xref="PID:g2996224"
                     /db_xref="GI:2996224"
                     /translation="MTVTLNRGSITSLMSSSQAVSTLQPVASELKTQLENKLKSESAE
                     KTREVLWQQYYASNPPDHAVLEVLATPVREALLARFGQHQGSVVPAIDLPELRSVLQQ
                     FDSFGKRWEAILLQVLEGIKPNESQVGLPYLSELINKELMILLPSNSIVDSLLHNSHQ
                     IDMDT"
```

```
     gene            complement(6588. .6935)
                     /gene="yscG"
     CDS             complement(6588. .6935)
                     /gene="yscG"
                     /codon_start=1
                     /transl_table=11
                     /product="Yop proteins translocation protein G homolog"
                     /protein_id="AAC62548.1"
                     /db_xref="PID:g2996225"
                     /db_xref="GI:2996225"
                     /translation="MKYKLNVLLAEIALIGTGNHYHEEANCIAEWLHLKGEEEAVQLI
                     RLSSLMNRGDYASALQQGNKLAYPDLEPWLALCEYRLGLGSALESRLNRLARSQDPRI
                     QTFVNGMREQLKT"
     gene            complement(6937. .7200)
                     /gene="yscF"
     CDS             complement(6937. .7200)
                     /gene="yscF"
                     /codon_start=1
                     /transl_table=11
                     /product="Yop proteins translocation protein F homolog"
                     /protein_id="AAC62549.1"
                     /db_xref="PID:g2996226"
                     /db_xref="GI:2996226"
                     /translation="MSNFSGFTKGTDIADLDAVAQTLKKPADDANKAVNDSIAALKDK
                     PDNPALLADLQHSINKWSVIYNINSTIVRSMKDLMQGILQKFP"
     gene            complement(7201. .7401)
                     /gene="yscE"
     CDS             complement(7201. .7401)
                     /gene="yscE"
                     /codon_start=1
                     /transl_table=11
                     /product="Yop proteins translocation protein E homolog"
                     /protein_id="AAC62550.1"
                     /db_xref="PID:g2996227"
                     /db_xref="GI:2996227"
                     /translation="MTQLEEQLHNVETVRSITMQLEMALTKLKKDMMRGGDAKQYQVW
                     QRESKALESAIAIIHYVAGDLK"
     gene            complement(7398. .8657)

alignment_scores:
            Quality:  202.50            Length:    200
              Ratio:  1.489               Gaps:      7
 Percent Similarity:  68.000   Percent Identity: 29.500

alignment_block:
 US-09-068-804-12 x AF053946   ..

 Align seg 1/1  to: AF053946  from: 1  to: 70504

      29 IleArgLeuThrPheAspGlyLysProValAlaLeuAlaGlyIleArgTh 45
         :::|||      ::::::::::::   :::   ::::::::::::|||||
   68960 GTACGCGCCGATCTTAATGCCAATTACATCCAGGTCGGTAACACTCGTAC 69009

      45 r.........GlnLysIleArgProAspArgLeuGluAlaHisMetLysM 59
         |         |||       ::::::::::|||||:::|||   :::|
   69010 CATAGCGTGCCAGTATCCGCTACAATCTCAACTTGAAAGCCATTTCCGTA 69059

      59 etLeuLeuGluLysGluCysSerCysLeuValValLeuThrSerGluArg 75
         |||||   |||:::      :::   |||:::|||||:::|||
   69060 TGCTGGCAGAAAACCGAACGCCAGTGTTGGCTGTTTTAGCGTCCAGTTCT 69109

      76 SerAspAlaGlyLys......ThrIleThrThrTyrPheArgGlySerTy 90
            |||::::::         ::::::   |||||||||   |||
   69110 GAGATAGCCAATCAAAGATTCGGTATGCCAGATTATTTCCGCCAGAGTGG 69159

      90 rThrPheGlyGluValHisThrAsnSerGlnLysValSerSerAlaSerG 107
        |||:::|||   :::   :::::::|||:::::::::      ::::::
   69160 TACCTATGGCAGTATCACTGTAGAGTCTAAAATGACTCAGCAAGTTGGTC 69209

     107 lnGlyGluAlaIle......AspGlnTyrAsnMetGlnLeu...SerCys 120
           |||::::::|||       |||   |||::::::   :::
   69210 TCGGTGACGGGATTATGGCAGATATGTATACTTTAACGATTCGTGAAGCG 69259

     121 GlyGluLysArgTyrThrIleProValLeuHisValLysAsnTrpProAs 137
         |||:::|||      ::::::|||||||:::||||||   ||||||||||
   69260 GGTCAAAAAACAATCTCTGTTCCTGTGGTTCATGTTGGCAATTGGCCCGA 69309

     137 pHisGlnProLeuProSer......ThrAspGlnLeuGluTyrLeuAlaA 152
         |:::   ::::::::::|||      |||      |||:::   |||:::|
   69310 TCAGACCGCAGTCAGCTCTGAAGTTACCAAGGCACTCGCTTCACTGGTAG 69359

     152 spArgValLysAsnSerAsnGlnAsn.............GlyAlaPro 163
         ||::::::   ::::::::::::|||                |||::::::
   69360 ATCAAACAGCAGAAACAAAACGCAATATGTATGAAAGCAAAGGAAGTTCA 69409

     164 GlyArgSerSerSerAspLysHisLeuProMetIleHisCysLeuGlyGl 180
         :::   :::::::::::|||      |||:::||||||||   :::||
   69410 GCGGTAGGAGATGACTCCAAATTACGGCCGGTAATACATTGCCGTGCGGG 69459

     180 yValGlyArgThrGlyThrMetAlaAlaAlaLeuValLeuLysAspAsnP 197
         |||||||||||||:::   :::   :::|||:::   ::::::|||:::
   69460 TGTTGGCCGTACTGCGCAACTGATTGGCGCAATGTGCATGAATGATAGTC 69509

     197 roHisSerAsnLeu......GluGlnValArgAlaAspPheArgIleHis 211
           :::|||:::|||      |||:::::::   ::::::   |||::::::
   69510 GTAATAGTCAGTTAAGCGTAGAAGATATGGTCAGCCAAATGCGAGTACAA 69559

seq_name: gb_ba2:AF074612

seq_documentation_block:
LOCUS       AF074612    70559 bp    DNA   circular  BCT       03-NOV-1998
DEFINITION  Yersinia pestis plasmid pCD1, complete plasmid sequence.
ACCESSION   AF074612
NID         g3822037
VERSION     AF074612.1  GI:3822037
KEYWORDS    .
SOURCE      Yersinia pestis.
  ORGANISM  Yersinia pestis
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Yersinia.
REFERENCE   1  (bases 1 to 70559)
  AUTHORS   Perry,R.D., Straley,S.C., Fetherston,J.D., Rose,D.J., Gregor,J. and
            Blattner,F.R.
  TITLE     DNA sequencing and analysis of the low-Ca2+-response plasmid pCD1
            of Yersinia pestis KIM5
  JOURNAL   Infect. Immun. 66 (10), 4611-4623 (1998)
  MEDLINE   98427122
REFERENCE   2  (bases 1 to 70559)
  AUTHORS   Perry,R.D., Straley,S.C., Fetherston,J.D., Rose,D.J., Gregor,J. and
            Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUN-1998) Microbiology and Immunology, University of
            Kentucky, MS415 Medical Center, Lexington, KY 40536-0084, USA
COMMENT     Unique identifiers, starting at Y0001, are assigned to each gene of
            pCD1.
FEATURES             Location/Qualifiers
     source          1. .70559
                     /organism="Yersinia pestis"
                     /plasmid="pCD1"
                     /strain="KIM5"
                     /db_xref="taxon:632"
     gene            57. .368
                     /gene="Y0001"
     CDS             57. .368
                     /gene="Y0001"
                     /note="o103; 43 pct identical (0 gaps) to 100 residues of
                     an approx. 200 aa protein GENPEPT: gi|537126, orf_o198
                     Escherichia coli"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAC69758.1"
                     /db_xref="PID:g3822038"
                     /db_xref="GI:3822038"
                     /translation="MHQQSRGAAGSRTLSLLMRQSGYNVVRWLARRLMRECGLASRQP
                     GKPRYRGEREVSLASPDLLKRQFKPSEPNRVWSGYISYIKVNGGWCYLALVIDLYSFH
```

```
                W"
gene            665. .1033
                /gene="nuc"
                /note="Y0002"
CDS             665. .1033
                /gene="nuc"
                /codon_start=1
                /transl_table=11
                /product="endonuclease"
                /protein_id="AAC69759.1"
                /db_xref="PID:g3822039"
                /db_xref="GI:3822039"
                /translation="MDTKLLQHTPIGTMVDYRPVNTKSGGKRLRRCPDFVIHYRMDLL
                VNAGIPVRTVNSFKALHDKVIIVDGKNTQMGSFNFSQAAVQSNSENVLIIWGDFTVVQ
                AYLQYWQSRWNKGTDWRSSY"
gene            1171. .1425
                /gene="repB"
                /note="Y0003"
CDS             1171. .1425
                /gene="repB"
                /codon_start=1
                /transl_table=11
                /product="replication protein"
                /protein_id="AAC69760.1"
                /db_xref="PID:g3822040"
                /db_xref="GI:3822040"
                /translation="MSQIENAVTSSSKRAYRKGNPLTGAEKQRMSVSRKKETHKAINV
                FIQNDLKNELLQLCEDSGLTQTEMIERWIQREKAARTNAA"
misc_RNA        complement(1560. .1649)
                /note="antisense RNA"
                /product="copA"
gene            1667. .1741
                /gene="tap"
                /note="Y0004"
CDS             1667. .1741
                /gene="tap"
                /codon_start=1
                /transl_table=11
                /product="repA translation protein"
                /protein_id="AAC69761.1"
                /db_xref="PID:g3822041"
                /db_xref="GI:3822041"
                /translation="MFRKVQYLLRLLLLPCNISAGRCD"
gene            1734. .2600
                /gene="repA"
                /note="Y0005"
CDS             1734. .2600
                /gene="repA"
                /codon_start=1
                /transl_table=11
                /product="replication protein"
                /protein_id="AAC69762.1"
                /db_xref="PID:g3822042"
                /db_xref="GI:3822042"
                /translation="MTNHQALFTHHYRQVKNPNPEFTPREGKKTLPFCRKLMAKAEGF
                TSRFDFSMHVAFARSLSLRHRMPPLLRRRAIDALLQGMCFHYDPLANRIQRSITNLAI
                ECGLATESKSGNLSITRATRALRFLSELGLITYQTEYDPQIGCNIPTDITFTPALFSA
                LDVSDVAVAAARRSRVEWENQQREKQRLPRLEMDELIAKAWRFVRERFRSYQTERKAH
                GLKRARARRDVDRTRRDIEAIVNRQLTREIAEGRFVGNLDAVRREKARRVKERMLMSR
                NNNYTRLATGAT"
rep_origin      1741. .2935
                /note="oriR"
gene            complement(3427. .3645)
                /gene="Y0006"
CDS             complement(3427. .3645)
                /gene="Y0006"
                /note="Orf5 (f72); 42 pct identical (0 gaps) to 33
                residues of an approx. 216 aa protein GENPEPT: gi|2055297,
                lmp2 Xenopus laevis"
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AAC69763.1"
                /db_xref="PID:g3822043"
                /db_xref="GI:3822043"
                /translation="MRSPLAYGCSVYTVNVVTQLHNDIHKSAYKHKRLYHVLTGQASS
                AQPVFSMPHRGMKAPADINYNSVDNWRR"
gene            4758. .5186
                /gene="Y0008"
CDS             4758. .5186
                /gene="Y0008"
                /note="Orf7 (o142); 31 pct identical (1 gap) to 48
                residues of an approx. 104 aa protein GENPEPT: gi|2149940,
                orf1 P. syringae"
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AAC69764.1"
                /db_xref="PID:g3822044"
                /db_xref="GI:3822044"
                /translation="MINTTFTELLPKIASHFGLDKLSQDEYGLCELILNDRVVIMLRA
                DEILNRLTLLGPILGFSGPEARSAASQLFFCYSINALNKDGPCFAWSEELGLIAFKHL
                SLDELNVENVSKEIANFYDWLSLVSLPAETAPSYSIYSIG"
gene            5204. .7402
                /gene="ypkA"
                /note="Y0009"
CDS             5204. .7402
                /gene="ypkA"
                /codon_start=1
                /transl_table=11
                /product="secreted protein kinase"
                /protein_id="AAC69765.1"
                /db_xref="PID:g3822045"
                /db_xref="GI:3822045"
                /translation="MKSVKIMGTMPPSISLAKAHERISQHWQNPVGELNIGGKRYRII
                DNQVLRLNPHSGFSLFREGVGKIFSGKMFNFSIARNLTDTLHAAQKTISQELRSDIPN
                ALSNLFGAKPQTELPLGWKGEPLSGAPDLEGMRVAETDKFAEGESHISIIETKDKQRL
                VAKIERSIAEGHLFAELEAYKHIYKTAGKHPNLANVHGMAVVPYGNRKEEALLMDEVD
                GWRCSDTLRTLADSWKQGKINSEAYWGTIKFIAHRLLDVTNHLAKAGVVHNDIKPGNV
                VFDRASGEPVVIDLGLHSRSGEQPKGFTESFKAPELGVGNLGASEKSDVFLVVSTLLH
                CIEGFEKNPEIKPNQGLRFITSEPAHVMDENGYPIHRPGIAGVETAYTRFITDILGVS
                ADSRPDSNEARLHEFLSDGTIDEESAKQILKDTLTGEMSPLSTDVRRITPKKLRELSD
                LLRTHLSSAATKQLDMGGVLSDLDTMLVALDKAEREGGVDKDQLKSFNSLILKTYRVI
                EDYVKGREGDTKNSSTEVSPYHRSNFMLSIVEPSLQRIQKHLDQTHSFSDIGSLVRAH
                KHLETLLEVLVTLSQQGQPVSSETYGFLNRLTEAKITLSQQLNTLQQQQESAKAQLSI
                LINRSGSWADVARQSLQRFDSTQPVVKFGTEQYTAIHRQMMAAHAAITLQEVSEFTDD
                MRNFTVDSIPLLIQLGRSSLMDEHLVEQREKLRELTTIAERLNRLEREWM"
gene            7798. .8664
                /gene="yopJ"
                /note="Y0010"
CDS             7798. .8664
                /gene="yopJ"
                /codon_start=1
                /transl_table=11
                /product="targeted effector protein"
                /protein_id="AAC69766.1"
                /db_xref="PID:g3822046"
                /db_xref="GI:3822046"
                /translation="MIGPISQINISGGLSEKETSSLISNEELKNIITQLETDISDGSW
                FHKNYSRMDVEVMPALVIQANNKYPEMNLNLVTSPLDLSIEIKNVIENGVRSSRFIIN
                MGEGGIHFSVIDYKHINGKTSLILFEPANFNSMGPAMLAIRIKTAIERYQLPDCHFSM
                VEMDIQRSSSECGIFSLALAKKLYIERDSLLKIHEDNIKGILSDGENPLPHDKLDPYL
                PVTFYKHTQGKKRLNEYLNTNPQGVGTVVNKKNETIVNRFDNNKSIVDGKELSVSVHK
                KRIAEYKTLLKV"
gene            9858. .10064
                /gene="Y0012"
CDS             9858. .10064
                /gene="Y0012"
                /note="o68; 45 pct identical (0 gaps) to 68 residues of an
                approx. 560 aa protein GENPEPT: gi|2182435, hypothetical"
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AAC69767.1"
                /db_xref="PID:g3822047"
                /db_xref="GI:3822047"
```

```
                  /translation="MAVGRKNWLFAGSLRAGQRMASILSLLETAKLNGHDPYVWLRDV
                  LTRLPTWPNSQLNALLPYAENRFS"
     gene         10347. .11753
                  /gene="yopH"
                  /note="Y0013"
     CDS          10347. .11753
                  /gene="yopH"
                  /codon_start=1

alignment_scores:
            Quality:  202.50            Length:    200
              Ratio:  1.489               Gaps:      7
 Percent Similarity:  68.000   Percent Identity: 29.500

alignment_block:
 US-09-068-804-12 x AF074612   ..

 Align seg 1/1  to: AF074612  from: 1  to: 70559

      29 IleArgLeuThrPheAspGlyLysProValAlaLeuAlaGlyIleArgTh 45
         :::|||      ::::::::::::   :::   ::::::::::::|||||
   11064 GTACGCGCCGATCTTAATGCCAATTACATCCAGGTCGGTAACACTCGTAC 11113

      45 r.........GlnLysIleArgProAspArgLeuGluAlaHisMetLysM 59
         |         |||          ::::::::::|||||:::|||   :::|
   11114 CATAGCGTGCCAGTATCCGCTACAATCTCAACTTGAAAGCCATTTCCGTA 11163

      59 etLeuLeuGluLysGluCysSerCysLeuValValLeuThrSerGluArg 75
         |||||   |||:::      :::   |||:::|||||:::|||
   11164 TGCTGGCAGAAAACCGAACGCCAGTGTTGGCTGTTTTAGCGTCCAGTTCT 11213

      76 SerAspAlaGlyLys......ThrIleThrThrTyrPheArgGlySerTy 90
            |||::::::         ::::::   |||||||||   |||
   11214 GAGATAGCCAATCAAAGATTCGGTATGCCAGATTATTTCCGCCAGAGTGG 11263

      90 rThrPheGlyGluValHisThrAsnSerGlnLysValSerSerAlaSerG 107
          |||:::|||   :::   ::::::|||:::::::::      ::::::
   11264 TACCTATGGCAGTATCACTGTAGAGTCTAAAATGACTCAGCAAGTTGGTC 11313

     107 lnGlyGluAlaIle......AspGlnTyrAsnMetGlnLeu...SerCys 120
           |||::::::|||      |||   |||::::::   :::
   11314 TCGGTGACGGGATTATGGCAGATATGTATACTTTAACGATTCGTGAAGCG 11363

     121 GlyGluLysArgTyrThrIleProValLeuHisValLysAsnTrpProAs 137
         |||:::|||      ::::::::::|||::::::::   ||||||||||
   11364 GGTCAAAAAACAATCTCIGTTCCTGTGGTTCATGTTGGCAATTGGCCCGA 11413

     137 pHisGlnProLeuProSer......ThrAspGlnLeuGluTyrLeuAlaA 152
         |:::   :::::::::|||      |||      |||:::   |||:::|
   11414 TCAGACCGCAGTCAGCTCTGAAGTTACCAAGGCACTCGCTTCACTGGTAG 11463

     152 spArgValLysAsnSerAsnGlnAsn..............GlyAlaPro 163
         ||:::::   ::::::::::::|||              |||::::::
   11464 ATCAAACAGCAGAAACAAAACGCAATATGTATGAAAGCAAAGGAAGTTCA 11513

     164 GlyArgSerSerSerAspLysHisLeuProMetIleHisCysLeuGlyGl 180
         :::   ::::::::::::|||      |||:::|||||||||   :::||
   11514 GCGGTAGGAGATGACTCCAAATTACGGCCGGTAATACATTGCCGTGCGGG 11563

     180 yValGlyArgThrGlyThrMetAlaAlaAlaLeuValLeuLysAspAsnP 197
         ||||||||||||:::   :::   :::|||:::   ::::::|||:::
   11564 TGTTGGCCGTACTGCGCAACTGATTGGCGCAATGTGCATGAATGATAGTC 11613

     197 roHisSerAsnLeu......GluGlnValArgAlaAspPheArgIleHis 211
         :::|||:::|||      \\\:::::::   ::::::   |||::::::
   11614 GTAATAGTCAGTTAAGCGTAGAAGATATGGTCAGCCAAATGCGAGTACAA 11663

seq_name: gb_pat:A52315

seq_documentation_block:
LOCUS       A52315          2405 bp    DNA             PAT       12-DEC-1997
DEFINITION  Sequence 19 from Patent EP0721989.
ACCESSION   A52315
```

```
NID         g2851996
VERSION     A52315.1  GI:2851996
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 2405)
  AUTHORS   Popoff,M.Y. and Le,G.F.
  TITLE     Oligonucleotides for the detection of salmonella
  JOURNAL   Patent: EP 0721989-A 19 17-JUL-1996;
            PASTEUR INSTITUT (FR)
COMMENT     Other publication JP 8317798 961203
            Other publication CA 2167354 960717
            Other publication FR 2729392 960719.
FEATURES             Location/Qualifiers
     source          1. .2405
                     /organism="unidentified"
                     /db_xref="taxon:32644"
     CDS             97. .1758
                     /note="GENE IAGA"
                     /codon_start=1
                     /protein_id="CAA03394.1"
                     /db_xref="PID:e1215753"
                     /db_xref="PID:g2851997"
                     /db_xref="GI:2851997"
                     /translation="MPHFNPVPVSNKKFVFDDFILNMDGSLVRSEKKVNIPPKEYAVL
                     VILLEAAGKIVSKNTLLDQVWGDAEVNEESLTRCIYALRRILSEDKEHRYIETLYGQG
                     YRFNRPVVVVSPPAPQPTTHTLAILPFQMQDQVQSESLHYSIVKGLSQYAPFGLSVLP
                     VTITKNCRSVKDILELMDQLRPDYYISGQMIPDGNDNIVQIEIVRVKGYHLLHQESIK
                     LIEHQPASLLQNKIANLLLRCIPGLRWDTKQISELNSIDSTMVYLRGKHELNQYTPYS
                     LQQALKLLTQCVNMSPNSIAPYCALAECYLSMAQMGIFDKQNAMIKAKEHAIKATELD
                     HNNPQALGLLGLINTIHSEYIVGSLLFKQANLLSPISADIKYYYGWNLFMAGQLEEAL
                     QTINECLKLDPTRAAAGITKLWITYYHTGIDDAIRLGDELRSQHLQDNPILLSMQVMF
                     LSLKGKHELARKLTKEISTQEITGLIAVNLLYAEYCQNSERALPTIREFLESEQRIDN
                     NPGLLPLVLVAHGEAIAEKMWNKFKNEDNIWFKRWKQDPRLIKLR"
     CDS             1776. .2258
                     /note="SEQUENCE DU GENE IAGB"
                     /codon_start=1
                     /protein_id="CAA03395.1"
                     /db_xref="PID:e1215754"
                     /db_xref="PID:g2851998"
                     /db_xref="GI:2851998"
                     /translation="MHYFFIIVIWLLSINTAWADCWLQAEKMFNIESELLYAIAQQES
                     AMKPGAIGHNRDGSTDLGLMQINSFHMKRLKKMGISEKQLLQDPCISVIVGASILSDM
                     MKIYGFSWEAVGAYNAGTSPKRSDIRKRYAKKIWENYRKLKEMSAEEKNKRLSIAVNK
                     "
BASE COUNT      735 a    488 c    519 g    663 t
ORIGIN


alignment_scores:
            Quality:  123.00            Length:     30
              Ratio:  4.241               Gaps:      0
 Percent Similarity:  96.667   Percent Identity: 96.667

alignment_block:
 US-09-068-804-12 x A52315/rev   ..

 Align seg 1/1  to reverse of: A52315  from: 1  to: 2405

     210 IleHisGlyThrIleGluCysTrpLysThrProAlaGlnPheValGlnLe 226
         |||||||||||||||||||||||||||||||||   ||||||||||||||
    2403 ATTCACGGAACAATAGAATGCTGGAAGACGCCTCA.CAGTTTGTACAACT 2355

     226 uLysAlaMetGlnAlaGlnLeuLeuMetThrThrAlaSer 239
         |||||||||||||||||||||||||||||||||||||||
    2354 AAAAGCAATGCAAGCCCAGTTGCTTATGACGACGGCAAGC 2315

seq_name: gb_bal:STIAGAB

seq_documentation_block:
LOCUS       STIAGAB      2406 bp    DNA             BCT       28-JUN-1995
DEFINITION  S.typhi genes iagA and iagB.
```

```
ACCESSION   X80892
NID         g791139
VERSION     X80892.1  GI:791139
KEYWORDS    .
SOURCE      Salmonella typhi.
  ORGANISM  Salmonella typhi
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Salmonella.
REFERENCE   1  (bases 1 to 2406)
  AUTHORS   Miras,I., Hermant,D., Arricau,N. and Popoff,M.Y.
  TITLE     Nucleotide sequence of iagA and iagB genes involved in invasion of
            HeLa cells by Salmonella enterica subsp. enterica ser. Typhi
  JOURNAL   Res. Microbiol. 146 (1), 17-20 (1995)
  MEDLINE   95273801
REFERENCE   2  (bases 1 to 2406)
  AUTHORS   Popoff,M.Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-AUG-1994) M.Y. Popoff, Institut Pasteur, Unite des
            Enterobacteries, 28 rue du Docteur Roux, 75724 Paris Cedex 15,
            FRANCE
REFERENCE   3  (bases 1 to 2406)
  AUTHORS   Chevrier,D., Popoff,M.Y., Dion,M.P., Hermant,D. and Guesdon,J.L.
  TITLE     Rapid detection of Salmonella subspecies I by PCR combined with
            non-radioactive hybridisation using covalently immobilised
            oligonucleotide on a microplate
  JOURNAL   FEMS Immunol. Med. Microbiol. 10 (3-4), 245-501 (1995)
  MEDLINE   95291265
FEATURES             Location/Qualifiers
     source          1. .2406
                     /organism="Salmonella typhi"
                     /strain="Ty2"
                     /db_xref="taxon:601"
     gene            98. .1759
                     /gene="iagA"
     CDS             98. .1759
                     /gene="iagA"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA56853.1"
                     /db_xref="PID:g791140"
                     /db_xref="GI:791140"
                     /db_xref="SWISS-PROT:P43016"
                     /translation="MPHFNPVPVSNKKFVFDDFILNMDGSLVRSEKKVNIPPKEYAVL
                     VILLEAAGKIVSKNTLLDQVWGDAEVNEESLTRCIYALRRILSEDKEHRYIETLYGQG
                     YRFNRPVVVVSPPAPQPTTHTLAILPFQMQDQVQSESLHYSIVKGLSQYAPFGLSVLP
                     VTITKNCRSVKDILELMDQLRPDYYISGQMIPDGNDNIVQIEIVRVKGYHLLHQESIK
                     LIEHQPASLLQNKIANLLLRCIPGLRWDTKQISELNSIDSTMVYLRGKHELNQYTPYS
                     LQQALKLLTQCVNMSPNSIAPYCALAECYLSMAQMGIFDKQNAMIKAKEHAIKATELD
                     HNNPQALGLLGLINTIHSEYIVGSLLFKQANLLSPISADIKYYYGWNLFMAGQLEEAL
                     QTINECLKLDPTRAAAGITKLWITYYHTGIDDAIRLGDELRSQHLQDNPILLSMQVMF
                     LSLKGKHELARKLTKEISTQEITGLIAVNLLYAEYCQNSERALPTIREFLESEQRIDN
                     NPGLLPLVLVAHGEAIAEKMWNKFKNEDNIWFKRWKQDPRLIKLR"
     gene            1777. .2259
                     /gene="iagB"
     CDS             1777. .2259
                     /gene="iagB"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA56854.1"
                     /db_xref="PID:g791141"
                     /db_xref="GI:791141"
                     /db_xref="SWISS-PROT:P43018"
                     /translation="MHYFFIIVIWLLSINTAWADCWLQAEKMFNIESELLYAIAQQES
                     AMKPGAIGHNRDGSTDLGLMQINSFHMKRLKKMGISEKQLLQDPCISVIVGASILSDM
                     MKIYGFSWEAVGAYNAGTSPKRSDIRKRYAKKIWENYRKLKEMSAEEKNKRLSIAVNK
                     "
BASE COUNT      736 a    488 c    519 g    663 t
ORIGIN


alignment_scores:
            Quality:  123.00            Length:     30
              Ratio:  4.241               Gaps:      0
 Percent Similarity:  96.667   Percent Identity: 96.667

alignment_block:
 US-09-068-804-12 x STIAGAB/rev   ..

 Align seg 1/1  to reverse of: STIAGAB  from: 1  to: 2406

     210 IleHisGlyThrIleGluCysTrpLysThrProAlaGlnPheValGlnLe 226
         ||||||||||||||||||||||||||||||||||   ||||||||||||||
    2404 ATTCACGGAACAATAGAATGCTGGAAGACGCCTCA.CAGTTTGTACAACT 2356

     226 uLysAlaMetGlnAlaGlnLeuLeuMetThrThrAlaSer 239
         |||||||||||||||||||||||||||||||||||||||
    2355 AAAAGCAATGCAAGCCCAGTTGCTTATGACGACGGCAAGC 2316

seq_name: gb_in1:DDIPTP1A

seq_documentation_block:
LOCUS       DDIPTP1A     1749 bp    DNA             INV       30-NOV-1992
DEFINITION  Dictyostelium discoideum phosphotyrosine phosphatase (PTP1) gene,
            complete cds.
ACCESSION   L07125
NID         g167861
VERSION     L07125.1  GI:167861
KEYWORDS    phosphotyrosine phosphatase.
SOURCE      Dictyostelium discoideum (strain AX-3) (library: 16 hr lambda ZAP
            of G.Schnitzler) slug DNA.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 1749)
  AUTHORS   Howard,P.K., Sefton,B.M. and Firtel,R.A.
  TITLE     Analysis of a spatially regulated phosphotyrosine phosphatase
            identifies tyrosine phosphorylation as a key regulatory pathway in
            Dictyostelium
  JOURNAL   Cell 71, 637-647 (1992)
  MEDLINE   93046662
FEATURES             Location/Qualifiers
     source          1. .1749
                     /organism="Dictyostelium discoideum"
                     /strain="AX-3"
                     /db_xref="taxon:44689"
                     /dev_stage="slug"
                     /tissue_lib="16 hr lambda ZAP of G.Schnitzler"
                     /map="II"
     gene            join(1. .306,407. .574,658. .1749)
                     /gene="PTP1"
     CDS             join(1. .306,407. .574,658. .1749)
                     /gene="PTP1"
                     /codon_start=1
                     /product="protein tyrosine phosphatase"
                     /protein_id="AAA33241.1"
                     /db_xref="PID:g167862"
                     /db_xref="GI:167862"
                     /translation="MGSVESSNQMNGSIENKTNKIDISVLRPLPTRSNSSISLSSSSH
                     SSFSRMGSLGSLPTNSGSSSPYYNNSSFDLVDEERIKSSIYNLKNHIKCIHKIKEEFR
                     LLEESVGPSETSEGDKKHNTSKNRYTNILPVNHTRVQLKKIQDKEGSDYINANYIDGA
                     YPKQFICTQGPLPNTIADFWRMVWENRCRIIVMLSRESENCRIKCDRYWPEQIGGEQF
                     SIYGNGNEVFGTYSVELVEVIQDPEREIITRNIRLTFEGETRDITQYQYEGWPDHNIP
                     DHTQPFRQLLHSITNRQNQIIPSSDRNVPIIVHCSAGVGRTGTFCTAVIMMKKLDHYF
                     KQLDATPIDQVVDPFTHLPITEYQSDNLDLKGLGYHFKSSIYNSNGINNNNNNNLNNN
                     NNINNNSNGSNNTPQTEPNNEEDDDDAAESTKYAIMDKYNSRIDFNLFSIVLKLREQR
                     PGMVQQLEQYLFCYKTILDEIYHRLNCKLGFSLPHVNNINNYNNYSNTTTTTTSSLAS
                     TTIIHPSTNSKLN"
     exon            <1. .306
                     /gene="PTP1"
                     /number=1
     intron          307. .406
                     /gene="PTP1"
                     /number=1
     exon            407. .574
                     /gene="PTP1"
                     /number=2
     misc_feature    450. .1590
```

```
                     /function="PTP conserved catalytic domain"
     intron          575. .657
                     /gene="PTP1"
                     /number=2
     exon            658. .>1749
                     /gene="PTP1"
                     /number=3
     misc_feature    1165. .1458
                     /gene="PTP1"
                     /note="region of insertion in PTP conserved domain"
BASE COUNT      697 a    228 c    241 g    583 t
ORIGIN


alignment_scores:
            Quality:  109.00              Length:    179
              Ratio:  1.058                 Gaps:      6
 Percent Similarity:  57.542   Percent Identity: 27.374

alignment_block:
 US-09-068-804-12 x DDIPTP1A   ..

 Align seg 1/1  to: DDIPTP1A  from: 1  to: 1749

      43 IleArgThrGlnLysIleArgProAspArgLeuGluAlaHisMetLysMe 59
         |||   ||||||         |||:::   :::::::::      :::|| 
     679 ATTTGTACACAAGGACCATTACCAAATACAATTGCTGATTTTTGGAGAAT 728

      59 tLeuLeuGluLysGluCysSerCysLeuValValLeuThrSerGluArgS 76
         |:::   |||:::   |||      :::|||:::|||:::   |||     
     729 GGTATGGGAAAATCGATGTCGTATTATTGTTATGCTTTCAAGAGAGTCTG 778

      76 erAsp.....................................AlaGly 79
           :::                                    :::||| 
     779 AAAACTGTAGAATAAAATGTGATAGATATTGGCCAGAACAAATTGGAGGT 828

      80 LysThrIleThrThrTyrPheArgGlySerTyrThrPheGlyGluValHi 96
            ::::::::::|||      |||:::   :::||||||          
     829 GAACAATTTTCAATATATGGTAATGGTAATGAAGTATTTGGCACATACAG 878

      96 sThrAsnSerGlnLysValSerSerAlaSerGlnGlyGluAlaIleAspG 113
          ::::::        |||      :::::::::   |||   |||   :  
     879 TGTCGAATTGGTTGAGGTTATCCAAGATCCAGAGAGAGAGATAATCACAA 928

     113 lnTyrAsnMetGlnLeuSerCys...GlyGluLysArgTyrThrIlePro 128
         ::   |||::::::|||:::     ||||||:::|||      |||:::  
     929 GA...AACATTAGACTTACATTTGAGGGTGAGACAAGA...GATATCACT 972

     129 ValLeuHisValLysAsnTrpProAspHisGlnProLeuProSerThrAs 145
            :::      :::|||||||||||||:::            |||::  
     973 CAATATCAATATGAAGGTTGGCCAGATCATAACATCCCAGATCACACTCA 1022

     145 pGlnLeuGluTyrLeuAlaAspArgValLysAsnSerAsnGlnAsnGlyA 162
         :::::::      |||   :::   ::::::|||   :::::::::     
    1023 ACCATTCCGTCAGTTGTTACATAGTATCACAAATCGTCAAAATCAAATCA 1072

     162 laProGlyArgSerSerSerAspLysHisLeuProMetIle...HisCys 177
         |||         ||||||||:::::::::|||:::|||   ||||||   
    1073 TACCA.........TCATCAGATAGAAATGTTCCAATCATCGTTCATTGT 1113

     178 LeuGlyGlyValGlyArgThrGlyThrMetAlaAlaAlaLeuValLeuLy 194
            :::||||||||||||||||||||      :::|||:::::::::::  
    1114 AGTGCAGGAGTTGGTAGAACCGGCACCTTTTGTACAGCCGTCATTATGAT 1163

     194 sAspAsnProHisSerAsnLeuGluGlnValArgAla 206
         :   :::   :::      :::   |||:::   |||              
    1164 GAAAAAGTTGGACCATTACTTTAAACAATTGGATGCT 1200

 seq_name: gb_in1:AF017083

 seq_documentation_block:
 LOCUS       AF017083     6227 bp    mRNA            INV       18-DEC-1997
 DEFINITION  Hirudo medicinalis receptor tyrosine phosphatase (HmLAR2) mRNA,
             complete cds.
 ACCESSION   AF017083
 NID         g2695654
 VERSION     AF017083.1  GI:2695654
 KEYWORDS    .
 SOURCE      medicinal leech.
   ORGANISM  Hirudo medicinalis
             Eukaryota; Metazoa; Annelida; Clitellata; Hirudinida; Hirudinea;
             Arynchobdellida; Hirudiniformes; Hirudinidae; Hirudo.
 REFERENCE   1  (bases 1 to 6227)
   AUTHORS   Gershon,T.R., Baker,M., Nitabach,M., Wu,P. and Macagno,E.R.
   TITLE     Two receptor tyrosine phosphatases of the LAR subfamily are
             expressed in neurons and muscle cells during development in leeches
   JOURNAL   Unpublished
 REFERENCE   2  (bases 1 to 6227)
   AUTHORS   Gershon,T.R., Baker,M., Nitabach,M., Wu,P. and Macagno,E.R.
   TITLE     Direct Submission
   JOURNAL   Submitted (01-AUG-1997) Biological Sciences, Columbia University,
             Columbia University, New York, NY 10027, USA
 REFERENCE   3  (bases 1 to 6227)
   AUTHORS   Gershon,T.R., Baker,M., Nitabach,M., Wu,P. and Macagno,E.R.
   TITLE     Direct Submission
   JOURNAL   Submitted (17-DEC-1997) Biological Sciences, Columbia University,
             Columbia University, New York, NY 10027, USA
   REMARK    Sequence update by submitter
 COMMENT     On Dec 18, 1997 this sequence version replaced gi:2564494.
 FEATURES             Location/Qualifiers
      source          1. .6227
                      /organism="Hirudo medicinalis"
                      /db_xref="taxon:6421"
      gene            1. .6227
                      /gene="HmLAR2"
      CDS             1. .6156
                      /gene="HmLAR2"
                      /note="similar to to LAR"
                      /codon_start=1
                      /product="receptor tyrosine phosphatase"
                      /protein_id="AAB91460.1"
                      /db_xref="PID:g2695655"
                      /db_xref="GI:2695655"
                      /translation="MKMLPTSATLARLLAFKFGQSPGSLIMIHRKSRPPRVVQKVMEG
                      NMATFWCGVVGKPAPEMTWYHMNRKLESFTYQQSSRKIFTHHNLSVLRVEPVKKRDEG
                      RYECEAHNGVGVPDRVSFNLSVYTKNDPIPAGFPEILVHPKLGSVERYRSTVMACSGK
                      GNPEPTFSWLKDFIPIEFTPVRFSVMPTGALQLLNSTYSYEGKYECIAENSHGVTNSQ
                      QATLVLKARRIPPHFSALPDNAEVVHGGSLNLTCIASGSPEPFIMWKRNNLDLSHRSN
                      NHSHGTGVLRLEGITESANYTCHAISELGTASHVVQVKVNVLPKPPSSLWITEVSPNT
                      VHLKWSPGNSDPVDSYIIRFRPRYSHPDNFTEFTDVGSDTTDHHIARLIRYTEYEFRV
                      YAVNKLGRSASSTPVDVVTGELVPGSEPPNVRARPVSGTTVVVTMGRTNHCATGSYKG
                      YKVYYTGQPVVTPVSLWTLHIVENSHLTTLTNLKENHTYTISVLAFTSVGDGPLSDNV
                      QVFTRPGVPDQPTNFLGESVSSTSIHIQWDRPASNNIVRYKLRYMDSNTRSNFSVDIG
                      RKRQDQRIIWYLLTDLAASTVYHMTLSAVAEGGEGPPTTTIQLQTHPYVLKSSPTFVS
                      AKALNSTCVEVTWLFPPSSSSSLPSLPPPSSSSSSSAIVDGFNLYFSETKNFDDKSPH
                      RSLNLTSDLRSATLHDLKKFTAYKIWVAGYNRAGEGPLSDVIIVKTEEDIPGEPKRVK
                      VEALNSTSVHLEWKAPQDGEHNGDIKGYVVFYTELDENEEEKFEPVRSDPLSSHQTEL
                      TIGGLSPSTLYQFQVCAYTRKGDGDKSKSKKVKTKAAVPSPPKGFTLSLMKEENPIVL
                      VTWQDPKFSHGVIEGYLVAFGVKGESHVEERRFNGNIHRFITPFLEKGATYEFKIAAK
                      NAVGYGEYATEVITTPDGAPSGAPQNLSAAVLGAHTIRLHWDPPVKKLRHGVIQMYEV
                      VYCESDDPGQDVMVNTTETSLVVDKLEADRSYTFIVRAFTSAGGSPWTNRVIKATSSS
                      TALPAPLNIRLRQTSVTSMEVSWDELHGKNPLNVLGYRIYYSSTPNANLDKWSIIDIG
                      PGTSAEVHGLDPSTTYSVRIRARGSDGRFGILSDQSSPTDWIMKIRKQFQVSSAELTT
                      PGPSYSSGRPPRKPGVLRYKLEYMGVRKTRDFSGMEDTTVDPRKEVTLQKEERSRSLD
                      ELMSRTLYTFNISARFIDGTWGPPTTIHVETNSEETILDPPSFIRPLSDQMVSLRVKP
                      AADVLGKVLQYLIFVVPTEVARTYEFLKNSISKNTQKPRCRGDGLQVGLMDQTFPRRL
                      SWGMAGEPMTFSTDPFPKGHNYVAVLRAVLGDNKVIFSPYSKPFFTDPAYLCSWYPTG
                      RVWIYLWIVAPICAAVVLILLLILVIIIARKRRVQDKNIAAPTNKVLLPAGSEMHPSF
                      VDPVDVRRMQFQSPAMVNHPPIRVEDLATHLEALRASDNNRFSLEYESIDPGQQFTWE
                      ASNLEVNKSKNRYANVIAYDHSRVTLHPLPQASSTYHPPGQVVPGSDYINANFLDGFR
                      KQNAYIATQGPMAETSVDFWRMVWERQSPVIVMMTKLEERGRVKCDQYWPSRGSESYG
                      SLQVNLVDVIELATYTMRTFQMSSEFNSEKREVRHFQFTAWPDYGVPEHAAPLLLFHE
                      EGQVHDPPDSGPIVVHCSAGVGRTGVFVVLDSMLERIKHTGSVDIYGHVTCLRAQRNY
                      MVQTEDQYIFIHAAILEAVTSGNTEVPARNLFAHMQKLLEPLTITSQSGHSTTITGIE
                      AEFKKLSSGKTSLSSFASANLSCNKQKNRLVNVLPYETTRVCLQPIRGVDGSDYINAS
                      FIDGYRYRRAYIATQGPLPDTVEDFWRALWESNCNIIVMLTKLREMGREMCHQYWPSE
```

```
                    RSARYQYFVVDPLAEYNMPQYILREFKVTDARDGQSRTMRQFQLTDWPEQGVPSTGDG
                    FIDFIGQTHKTKEQFGQEGPIAVHCSAGVGRTGVFITLSIVLERMRFEGAVDVFQTVN
                    VLRTQRPGMVQTEEQYAFCYRAALEYLGSFDHYTN"
BASE COUNT      1702 a    1621 c    1467 g    1436 t      1 others
ORIGIN


alignment_scores:
           Quality:  102.50              Length:    228
             Ratio:  0.958                 Gaps:     12
 Percent Similarity:  46.930   Percent Identity: 22.807

alignment_block:
 US-09-068-804-12 x AF017083   ..

 Align seg 1/1  to: AF017083  from: 1  to: 6227

      10 HisSerIleAsnGlnGlnThrGln.ValLysLeuSerAspGlyMetProV 26
         :::|||:::   ||||||||||:::    :::|||   :::   :::|||
    5409 CAATCTGTCCTGCAACAAACAGAAAAACCGACTTGTCAATGTCTTGCCA. 5457

      26 alProValIleArgLeuThrPheAspGlyLysProValAlaLeuAlaGly 42
                               ::::::   :::   |||   |||
    5458 ....................TACGAAACGACTAGGGTCTGTTTGCAACCA 5487

      43 IleArg..................................................  44
         ||||||
    5488 ATCAGAGGAGTGGATGGATCAGATTATATAAACGCCAGTTTCATTGATGG 5537

      45 ............................ThrGlnLysIleArgProAspA 52
                                     ||||||      ||||||
    5538 TTACAGATACAGGAGGGCCTACATAGCCACACAGGGACCTCTGCCAGACA 5587

      52 rgLeuGluAlaHisMetLysMetLeuLeuGluLysGluCysSerCysLeu 68
           :::|||:::      :::   |||   |||   :::|||:::   :::
    5588 CAGTTGAGGATTTCTGGCGAGCTCTTTGGGAGAGCAACTGCAACATCATT 5637

      69 ValValLeuThrSerGluArgSerAspAlaGlyLysThrIleThrThrTy 85
         |||:::||||||      |||   :::   |||:::   :::
    5638 GTCATGCTCACCAAACTCAGA...GAAATGGGAAGGGAAATGTGTCATCA 5684

      85 rPheArgGlySer..............TyrThrPheGlyGluValHisT 97
           ::::::   |||              |||   :::      |||
    5685 ATACTGGCCAAGTGAGAGATCAGCAAGATATCAGTACTTCGTAGTT.... 5730

      97 hrAsnSerGlnLysValSerSerAlaSerGlnGlyGluAlaIleAspGln 113
                                          ::::::::::::::
    5731 ................................GACCCTTTAGCTGAG 5745

     114 TyrAsnMetGlnLeuSerCys............................. 120
         ||||||||||:::
    5746 TACAACATGCCCCAGTACATCCTCAGAGAGTTCAAGGTCACTGATGCAAG 5795

     121 ....GlyGluLysArgTyrThrIleProValLeuHisValLysAsnTrpP 136
             |||:::   |||   |||:::      ::::::::::::::||||
    5796 GGACGGACAGTCGAGG...ACAATGAGGCAGTTCCAATTGACAGACTGGC 5842

     136 roAspHisGlnProLeuProSerThrAspGln......LeuGluTyrLeu 150
         ||:::   |||   :::|||||||||:::::      ::::::::::::
    5843 CTGAA...CAGGGTGTCCCAAGTACTGGTGATGGTTTCATTGATTTCATT 5889

     151 AlaAspArgValLysAsnSerAsnGlnAsnGlyAlaProGlyArgSerSe 167
         ::::::      |||:::   :::|||   |||      |||      ::
    5890 GGGCAGACCCATAAGACCAAAGAACAATTTGGACAAGAGGGTCCTATAGC 5939

     167 rSerAspLysHisLeuProMetIleHisCysLeuGlyGlyValGlyArgT 184
         :                  :::||||||   :::|||||||||||||||
    5940 T.....................GTTCACTGCAGTGCTGGAGTTGGGAGAA 5968

     184 hrGlyThr...MetAlaAlaAlaLeuValLeu 193
         |||||:::   ::::::   ::::::|||||
    5969 CAGGAGTCTTCATAACTCTGAGCATCGTCTTG 6000
```

```
seq_name: gb_pr2:HSRPTYRPH

seq_documentation_block:
LOCUS       HSRPTYRPH    4389 bp    mRNA            PRI       14-JAN-1997
DEFINITION  H.sapiens mRNA for receptor protein tyrosine phosphatase.
ACCESSION   X95712
NID         g1666422
VERSION     X95712.1  GI:1666422
KEYWORDS    fmi gene; FMI protein; receptor protein tyrosine phosphatase.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4389)
  AUTHORS   Crossland,S., Smith,P.D. and Crompton,M.R.
  TITLE     Molecular cloning and characterization of PTP pi, a novel
            receptor-like protein-tyrosine phosphatase
  JOURNAL   Biochem. J. 319 (Pt 1), 249-254 (1996)
  MEDLINE   97024447
REFERENCE   2  (bases 1 to 4389)
  AUTHORS   Crossland,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-FEB-1996) S. Crossland, Institute of Cancer Research,
            Haddow Labs, 15 Cotswold Rd, Sutton, Surrey SM2 5NG, UK
FEATURES             Location/Qualifiers
     source          1. .4389
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="mammary"
                     /cell_type="carcinoma"
                     /cell_line="MCF-7"
                     /clone_lib="lambda ZAP II"
                     /clone="MCF7-fmi"
     gene            61. .4389
                     /gene="fmi"
     CDS             61. .4389
                     /gene="fmi"
                     /note="receptor protein tyrosine phosphatase"
                     /codon_start=1
                     /product="FMI protein"
                     /protein_id="CAA65016.1"
                     /db_xref="PID:e224170"
                     /db_xref="PID:g1666423"
                     /db_xref="GI:1666423"
                     /db_xref="SPTREMBL:Q92735"
                     /translation="MGGDAPTMARAQALVLELTFQLCAPETETPEVGCTFEEGSDPAV
                     PCEYSQAQYDDFQWDPGSPADLPHGSYLMVNTSQHAPGQRAHVIFQSLSENDTHCVQF
                     SYFLYSRDGHSPGTLGVYVRVNGGPLGSAVWNMTGSHGRQWHQAELAVSTFWPNEYQV
                     LFEALISPDRRGYMGLDDICILSYPCAKAPHFSRLGDVEVNAGQNASFQCMAAGRAEA
                     ERFLLQRQSGALVPAAGVRHISHRSFLATFPLAAVSRAEQDLYRCVSQAPRGRGTSLN
                     FAEFMVKEPPTPIAPPQLLRAGPTYLIIQLNTNSIIGDGPIVRKEIEYRMARGPWAEV
                     HAVSLQTYKLWHLDPDTEYEISVLLIRPGDGGTGRLGPPFISRTKCAEPMRAPKGLAF
                     AEIQARQLTLQWEPLGYNVTRCHTYTVSLCYHYTLGSSHNQTIPRVCEDRARCQPLHH
                     EEPAALSERSREVVLTNPEGRKEGKEVTFQTDEDVPSGIAAESLIFTPLEDMIFLKWE
                     EPQEPNGLITQYEISYQSIESSDPAVNVQATSTISKLRNETYHVFSNLHPGTTYLFSV
                     GARTGKGFGQAALTEITTYISAPSLDYADMPSPLGESENNITVLLRPAQGRGAPISVY
                     QVIVEEEQGSRRLRREPGGQDCFPVPLTFEAALARGLVDYFGAELAASSLPEAMPFTV
                     GDNKTYRGFWNPPLEPRKAYLIYFQAASHLKGETRLNCIRIARKAACKESKRPLEVSQ
                     RSEEMGLILGICAGGLAVLILLLGAIIVIIRKGRDHYAYSYYPKPVNMTKATVNYRQE
                     KTHMMSAVDRSFTDQSTLQPPGLSFMDTHGYSTRGEQRSGGVTEASSLLGGSPRRPCG
                     RKGSPYHTVQLHPAVRVADLLQHINQMKTAEGYGFKQEYESFFEGWDATKKKDKVKGS
                     RQEPMPAYDRHRVKLHPMLGDPNADYINANYIDGYHRSNHFIATQGPKPEMVYDFWRM
                     VWQEHCSSIVMITKLVEVAGCKCSRYWPEDSDTYGDIKIMLVKTEILAEYVVRTFALE
                     RRGYSARYEVRQFHFTAWPEHGVPYHTTGLLAFIRRVKASTPPDAGPIVIHCSAGTGR
                     TGCYIVLDVMLDMAECEGVVDIYNCVKTLCSRCVNMIQTEEQYIFIHDAILEACLCGE
                     TTIPVSEFKATYKEMIRIDPQSNSSQLREEFQTLNSVTPPLDVEEYSIALLPRNRDKN
                     RSMDVLPPDRCLPFLISTDGDSNNYINAALTDSYTRSAAFIVTLHPLQSTTPDFWRLV
                     YDYGCTSIVMLNQLNQSNSAWPCLQYWPEPGRQQYGLMEVEFMSGTADEDLVARVFRV
                     QNISRLQEGHLLVRHFQFLRWSAYRDTPDSEKAFLHLLAEVDKWQAESGDGRTIVHCL
                     NGGGRRGTSCALRTVLEMIRCHNLVDVSFAAKTLRNYKPNMVETMDQYHFCYDVALEY
                     LEAGVR"
     misc_feature    931. .2085
                     /gene="fmi"
```

```
                     /note="fibronectin repeats"
BASE COUNT      899 a    1382 c    1306 g     802 t
ORIGIN


alignment_scores:
            Quality:   99.00              Length:    150
              Ratio:  1.165                Gaps:      6
 Percent Similarity:  56.667   Percent Identity: 27.333

alignment_block:
 US-09-068-804-12 x HSRPTYRPH    ..

 Align seg 1/1  to: HSRPTYRPH  from: 1  to: 4389

      43 IleArgThrGlnLysIleArgProAspArgLeuGluAlaHisMetLysMe 59
         |||   ||||||      :::|||:::   :::   :::      :::||
    2926 ATAGCCACTCAAGGGCCGAAGCCTGAGATGGTCTATGACTTCTGGCGTAT 2975

      59 tLeuLeuGluLysGluCysSerCysLeuValValLeuThrSerGluArgS 76
         |:::   :::   :::||||||:::::::|||:::::|||
    2976 GGTGTGGCAGGAGCACTGTTCCAGCATCGTCATGATCACCAAGCTGGTCG 3025

      76 erAspAlaGlyLysThrIleThrThrTyrPheArg...GlySerTyrThr 91
             ||||||   :::   :::   |||         :::|||   |||
    3026 AGGTGGCAGGGTGTAAATGCTCACGGTACTGGCCGGAGGACTCAGACACC 3075

      92 PheGlyGluValHisThrAsnSerGlnLysValSerSerAlaSerGlnGl 108
         :::|||::::::   :::         |||:::
    3076 TACGGGGACATCAAGATTATGCTGGTGAAGACA..................  3108

     108 yGluAlaIleAspGlnTyrAsnMetGlnLeuSerCysGlyGluLys.... 123
          |||:::::::::::::|||   ::::::            |||:::
    3109 .GAGACCCTGGCTGAGTATGTCGTGCGCACTTTTGCCCTGGAGCGGAGAG 3157

     124 ..........ArgTyrThrIleProValLeuHisValLysAsnTrpPro 136
                 ||||||   :::      :::|||   ::::::|||||
    3158 GCTACTCTGCCCGGTACGAGGTCCGCCAGTTCCACTTCACAGCGTGGCCA 3207

     137 AspHis...GlnProLeuProSerThrAspGlnLeuGluTyrLeuAlaAs 152
         :::|||      |||      :::|||:::   |||:::::::::
    3208 GAGCATGGCGTCCCCTACCATACCACGGGGCTGCTGGCTTTCATC...CG 3254

     152 pArgValLysAsnSerAsnGlnAsnGlyAlaProGlyArgSerSerSerA 169
          |||||||||:::|||::::::   :::|||
    3255 GCGCGTGAAGGCCTCCACCCCACCTGATGCCGGGCCCATT.......... 3294

     169 spLysHisLeuProMetIleHisCysLeuGlyGlyValGlyArgThrGly 185
                   :::|||||||||   :::|||:::||||||||||||
    3295 ..............GTCATCCACTGCAGCGCGGGCACCGGCCGCACAGGT 3330

seq_name: gb_ro:RRU66566

seq_documentation_block:
LOCUS       RRU66566     1627 bp    mRNA            ROD       07-FEB-1997
DEFINITION  Rattus rattus rat receptor type protein tyrosine phophatase psi
            (RPTPpsi) mRNA, partial cds.
ACCESSION   U66566
NID         g1519395
VERSION     U66566.1  GI:1519395
KEYWORDS    .
SOURCE      black rat.
  ORGANISM  Rattus rattus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 1627)
  AUTHORS   Sommer,L., Rao,M. and Anderson,D.J.
  TITLE     RPTP delta and the novel protein tyrosine phosphatase RPTP psi are
            expressed in restricted regions of the developing central nervous
            system
  JOURNAL   Dev. Dyn. 208 (1), 48-61 (1997)
  MEDLINE   97143509
REFERENCE   2  (bases 1 to 1627)
  AUTHORS   Sommer,L., Rao,M. and Anderson,D.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-AUG-1996) Division of Biology, Howard Hughes Medical
            Institute, California Institute of Technology, Pasadena, CA 91125,
            USA
FEATURES             Location/Qualifiers
     source          1. .1627
                     /organism="Rattus rattus"
                     /db_xref="taxon:10117"
     gene            1. .1627
                     /gene="RPTPpsi"
     CDS             <1. .>1627
                     /gene="RPTPpsi"
                     /codon_start=2
                     /product="rat receptor type protein tyrosine phophatase
                     psi"
                     /protein_id="AAB42210.1"
                     /db_xref="PID:g1519396"
                     /db_xref="GI:1519396"
                     /translation="GESENTITVLLRPAQGRGAPISAYQVVVEEERPRRLRREPGGQD
                     CFSVPLTFETALARGLVHYFGAELAASSLLEAMPFTVGDNQTYRGFWNPPLEPRKAYL
                     IYFQAASHLEGETRLNCIRIARKAACKESKRPLEVSQRSEEMGLILGICAGGLVVLIL
                     LLGAIIIIIRKGRDRYAYSYYPKPVNMAKATVNYRQEKTHMMTAVDRSFTDQSTLQED
                     ERLGLSFMDAPGYSPRGDQRSSGVTEASSLLGGSPRRPCGRKGSPYHTGQLHPAVRVA
                     DLLQHINQMKTAEGYGFKQEYESFFEGWDATKKKDKLKSGRQEPVSAYDRHRVKLHPM
                     LGDPDADYISANYIDGYHRPNHFIATQGPKPEMIYDFWRMVWQEQCVSVVMISKLVEV
                     GRVKCSRYWPEHSDMYGDIKITLVKTETLAEYVVRTFALERRGYSARHEVRQFHFTAW
                     PEHGVPYHATGLLAFIRRVKASTPPDAGPVVIHCSAGTGRTGCYIVLDVMLDMAECEG
                     VVDIYNCVKTLCSRRVNMIQTEEQYIFIHDAILE"
BASE COUNT      354 a    464 c    495 g    314 t
ORIGIN


alignment_scores:
            Quality:   91.00              Length:    151
              Ratio:  1.046                Gaps:      7
 Percent Similarity:  57.616   Percent Identity: 23.841

alignment_block:
 US-09-068-804-12 x RRU66566    ..

 Align seg 1/1  to: RRU66566  from: 1  to: 1627

      43 IleArgThrGlnLysIleArgProAspArgLeuGluAlaHisMetLysMe 59
         |||   ||||||      :::|||:::   :::   :::      :::||
    1073 ATAGCCACTCAAGGGCCAAAGCCTGAGATGATCTACGACTTCTGGCGCAT 1122

      59 tLeuLeuGluLysGluCysSerCysLeuValValLeuThrSerGluArgS 76
         |:::   :::   :::|||   :::   ||||||::::::|||
    1123 GGTGTGGCAGGAGCAGTGTGTGAGC...GTCGTCATGATCAGCAAGCTGG 1169

      76 erAspAlaGlyLysThrIleThrThrTyrPheArgGlySerTyrThr... 91
           ::::::|||::::::      :::   ::::::      ::::::
    1170 TAGAGGTGGGCAGGGTAAAGTGCTCTCGATACTGGCCCGAGCACTCAGAC 1219

      92 ...PheGlyGluValHisThrAsnSerGlnLysValSerSerAlaSerGl 107
            :::|||::::::   ::::::      |||:::
    1220 ATGTATGGGGACATCAAGATCACGCTGGTAAAGACA.............. 1255

     107 nGlyGluAlaIleAspGlnTyrAsnMetGlnLeuSerCysGlyGluLys. 123
          |||:::::::::::|||   ::::::            |||:::
    1256 ....GAGACACTGGCTGAGTACGTGGTGCGCACCTTTGCCCTGGAGCGGA 1301

     124 ..............ArgTyrThrIleProValLeuHisValLysAsnTrp 135
                    |||:::   :::      :::|||   :::::|||
    1302 GAGGTTACTCAGCCCGGCATGAGGTCCGCCAGTTCCACTTCACAGCGTGG 1351

     136 ProAspHis...GlnProLeuProSerThrAspGlnLeuGluTyrLeuAl 151
         |||:::|||      |||      :::|||:::   |||::::::::::
    1352 CCAGAGCATGGTGTTCCCTACCATGCCACGGGGCTGCTGGCCTTCATC.. 1399

     151 aAspArgValLysAsnSerAsnGlnAsnGlyAlaProGlyArgSerSerS 168
          |||||||||:::|||::::::   :::|||
```

```
    1400 .CGGCGCGTGAAGGCTTCCACTCCACCTGATGCTGGGCCT..... 1438

     168 erAspLysHisLeuProMetIleHisCysLeuGlyGlyValGlyArgThr 184
                      :::|||||||||   :::|||:::|||||||||
    1439 .............GTTGTCATTCACTGCAGTGCAGGCACTGGCCGCACA 1474

     185 Gly 185
         |||
    1475 GGC 1477

seq_name: gb_pr2:HSLARR

seq_documentation_block:
LOCUS       HSLARR       7702 bp    mRNA            PRI       19-SEP-1995
DEFINITION  Human mRNA for LCA-homolog. LAR protein (leukocyte antigen
            related).
ACCESSION   Y00815
NID         g34266
VERSION     Y00815.1  GI:34266
KEYWORDS    antigen; cell surface glycoprotein; glycoprotein; immunoglobulin
            superfamily; LAR gene; leukocyte common antigen; neural cell
            adhesion molecule; transmembrane protein.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 7702)
  AUTHORS   Saito,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-SEP-1988) Saito H., Dana-Farber Cancer Institute, 44
            Binney Street, Boston, MA 02115
REFERENCE   2  (bases 1 to 7702)
  AUTHORS   Streuli,M., Krueger,N.X., Hall,L.R., Schlossman,S.F. and Saito,H.
  TITLE     A new member of the immunoglobulin superfamily that has a
            cytoplasmic region homologous to the leukocyte common antigen
  JOURNAL   J. Exp. Med. 168 (5), 1523-1530 (1988)
  MEDLINE   89035978
REFERENCE   3  (bases 1 to 7702)
  AUTHORS   Schaapveld,R.Q., van den Maagdenberg,A.M., Schepens,J.T.,
            Weghuis,D.O., Geurts van Kessel,A., Wieringa,B. and Hendriks,W.J.
  TITLE     The mouse gene Ptprf encoding the leukocyte common antigen-related
            molecule LAR: cloning, characterization, and chromosomal
            localization
  JOURNAL   Genomics 27 (1), 124-130 (1995)
  MEDLINE   95394448
FEATURES             Location/Qualifiers
     source          1. .7702
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="tonsil"
                     /cell_type="lymphocyte"
                     /clone_lib="lambda gt11"
     sig_peptide     371. .418
                     /product="put. signal peptide (AA -16 to -1)"
     CDS             371. .6064
                     /codon_start=1
                     /product="put. LAR preprotein (AA -16 to 1881)"
                     /protein_id="CAA68754.1"
                     /db_xref="PID:g34267"
                     /db_xref="GI:34267"
                     /db_xref="SWISS-PROT:P10586"
                     /translation="MVPLVPALVMLGLVAGAHGDSKPVFIKVPEDQTGLSGGVASFVC
                     QATGEPKPRITWMKKGKKVSSQRFEVIEFDDGAGSVLRIQPLRVQRDEAIYECTATNS
                     LGEINTSAKLSVLEEEQLPPGFPSIDMGPQLKVVEKARTATMLCAAGGNPDPEISWFK
                     DFLPVDPATSNGRIKQLRSGALQIESSEESDQGKYECVATNSAGTRYSAPANLYVRVR
                     RVAPRFSIPPSSQEVMPGGSVNLTCVAVGAPMPYVKWMMGAEELTKEDEMPVGRNVLE
                     LSNVVRSANYTCVAISSLGMIEATAQVTVKALPKPPIDLVVTETTATSVTLTWDSGNS
                     EPVTYYGIQYRAAGTEGPFQEVDGVATTRYSIGGLSPFSEYAFRVLAVNSIGRGPPSE
                     AVRARTGEQAPSSPPRRVQARMLSASTMLVQWEPPEEPNGLVRGYRVYYTPDSRRPPN
                     AWHKHNTDAGLLTTVGSLLPGITYSLRVLAFTAVGDGPPSPTIQVKTQQGVPAQPADF
                     QAEVESDTRIQLSWLLPPQERIIMYELVYWAAEDEDQQHKVTFDPTSSYTLEDLKPDT
                     LYRFQLAARSDMGVGVFTPTIEARTAQSTPSAPPQKVMCVSMGSTTVRVSWVPPPADS
                     RNGVITQYSVAHEAVDGEDRGRHVVDGISREHSSWDLVGLEKWTEYRVWVRAHTDVGP
                     GPESSPVLVRTDEDVPSGPPRKVEVEPLNSTAVHVYWKLPVPSKQHGQIRGYQVTYVR
                     LENGEPRGLPIIQDVMLAEAQWRPEESEDYETTISGLTPETTYSVTVAAYTTKGDGAR
                     SKPKIVTTTGAVPGRPTMMISTTAMNTALLQWHPPKELPGELLGYRLQYCRADEARPN
                     TIDFGKDDQHFTVTGLHKGTTYIFRLAAKNRAGLGEEFEKEIRTPEDLPSGFPQNLHV
                     TGLTTSTTELAWDPPVLAERNGRIISYTVVFRDINSQQELQNITTDTRFTLTGLKPDT
                     TYDIKVRAWTSKGSGPLSPSIQSRTMPVEQVFAKNFRVAAAMKTSVLLSWEVPDSYKS
                     AVPFKILYNGQSVEVDGHSMRKLIADLQPNTEYSFVLMNRGSSAGGLQHLVSIRTAPD
                     LLPHKPLPASAYIEDGRFDLSMPHVQDPSLVRWFYIVVVPIDRVGGSMLTPRWSTPEE
                     LELDELLEAIEQGGEEQRRRRRQAERLKPYVAAQLDVLPETFTLGDKKNYRGFYNRPL
                     SPDLSYQCFVLASLKEPMDQKRYASSPYSDEIVVQVTPAQQQEEPEMLWVTGPVLAVI
                     LIILIVIAILLFKRKRTHSPSSKDEQSIGLKDSLLAHSSDPVEMRRLNYQTPGMRDHP
                     PIPITDLADNIERLKANDGLKFSQEYESIDPGQQFTWENSNLEVNKPKNRYANVIAYD
                     HSRVILTSIDGVPGSDYINANYIDGYRKQNAYIATQGPLPETMGDFWRMVWEQRTATV
                     VMMTRLEEKSRVKCDQYWPARGTETCGLIQVTLLDTVELATYTVRTFALHKSGSSEKR
                     ELRQFQFMAWPDHGVPEYPTPILAFLRRVKACNPLDAGPMVVHCSAGVGRTGCFIVID
                     AMLERMKHEKTVDIYGHVTCMRSQRNYMVQTEDQYVFIHEALLEAATCGHTEVPARNL
                     YAHIQKLGQVPPGESVTAMELEFKLLASSKAHTSRFISANLPCNKFKNRLVNIMPYEL
                     TRVCLQPIRGVEGSDYINASFLDGYRQQKAYIATQGPLAESTEDFWRMLWEHNSTIIV
                     MLTKLREMGREKCHQYWPAERSARYQYFVVDPMAEYNMPQYILREFKVIDARDGQSRT
                     IRQFQFTDWPEQGVPKTGEGFIDFIGQVHKTKEQFGQDGPITVHCSAGVGRTGVFITL
                     SIVLERMRYEGVVDMFQTVKTLRTQRPAMVQTEDQYQLCYRAALEYLGSFDHYAT"
     mat_peptide     419. .6061
                     /product="put. LAR protein (AA 1 - 1881)"
     misc_feature    419. .4120
                     /note="put. extracellular domain"
     misc_feature    4121. .4192
                     /note="put. transmembrane domain"
     misc_feature    4193. .6061
                     /note="put. cytoplasmic domain"
BASE COUNT     1636 a   2286 c   2292 g   1488 t
ORIGIN


alignment_scores:
            Quality:   94.50            Length:    260
              Ratio:  0.768              Gaps:     13
 Percent Similarity:  47.308   Percent Identity: 21.154

alignment_block:
 US-09-068-804-12 x HSLARR   ..

 Align seg 1/1  to: HSLARR  from: 1  to: 7702

       8 GlnGlnHisSerIleAsnGlnGlnThrGln.ValLysLeuSerAspGlyM 24
         |||:::::::::      ||||||:::|||    :::|||   :::   |
    5311 CAGCGCCAACCTGCCCTGCAACAAGTTCAAGAACCGGCTGGTGAACATCA 5360

      24 etProValProValIleArgLeuThrPheAspGly.............. 35
         |||||      ::::::|||:::   ::::::
    5361 TGCCCTACGAATTGACCCGTGTGTGTCTGCAGCCCATCCGTGGTGTGGAG 5410

      36 .............LysProValAlaLeuAlaGlyIleArgThrGlnLy 47
                      ::::::      |||:::|||   |||   |||||
    5411 GGCTCTGACTACATCAATGCCAGCTTCCTGGATGGTTATAGACAGCAGAA 5460

      47 sIle.....................ArgProAspArgLeuGluAlaHisM 57
         |                        ::::::      |||:::
    5461 GGCCTACATAGCTACACAGGGGCCTCTGGCAGAGAGCACCGAGGACTTCT 5510

      57 etLysMetLeuLeuGluLysGluCysSerCysLeuValValLeuThrSer 73
           :::||||||   |||  :::::::::   :::|||:::||||||
    5511 GGCGCATGCTATGGGAGCACAATTCCACCATCATCGTCATGCTGACCAAG 5560

      74 GluArgSerAspAlaGlyLysThrIleThrThrTyrPhe.......... 86
           |||      :::      :::         |||
    5561 CTTCGGGAGATGGGCAGGGAGAAATGCCACCAGTACTGGCCAGCAGAGCG 5610

      87 .ArgGlySerTyrThrPheGlyGluValHisThrAsnSerGlnLysValS 103
          ::   |||   :::      |||
    5611 CTCTGCTCGCTACCAGTACTTTGTTGTT..................... 5638

     103 erSerAlaSerGlnGlyGluAlaIleAspGlnTyrAsnMetGlnLeuSer 119
                      ::::::::::::::|||||||||:::
```

```
  5639 ................GACCCGATGGCTGAGTACAACATGCCCCAGTAT 5671

   120 Cys................................GlyGluLysArgTy 125
                                          |||:::   |||
  5672 ATCCTGCGTGAGTTCAAGGTCACGGATGCCCGGGATGGGCAGTCAAGG.. 5719

   125 rThrIleProValLeuHisValLysAsnTrpProAspHisGlnProLeuP 142
        ||||||      ::::::   ::::::||||||:::   |||   :::|
  5720 .ACAATCCGGCAGTTCCAGTTCACAGACTGGCCAGAG...CAGGGCGTGC 5765

   142 roSerThrAspGln......LeuGluTyrLeuAlaAspArgValLysAsn 156
       ||   |||::::::      ::::::::::::::::::      |||:::
  5766 CCAAGACAGGCGAGGGATTCATTGACTTCATCGGGCAGGTGCATAAGACC 5815

   157 SerAsnGlnAsnGlyAlaProGlyArgSerSerSerAspLysHisLeuPr 173
          :::|||   |||      |||      :::
  5816 AAGGAGCAGTTTGGACAGGATGGGCCTATCACG..................  5848

   173 oMetIleHisCysLeuGlyGlyValGlyArgThrGlyThrMetAlaAlaA 190
          :::||||||   :::||||||||||||||||||:::      :::
  5849 ....GTGCACTGCAGTGCTGGCGTGGGCCGCACCGGGGTGTTCATCACTC 5894

   190 laLeuValLeuLysAspAsnProHisSerAsnLeuGluGlnValArgAla 206
          ::::::                        ||||||:::::|||
  5895 TGAGCATCGTC.....................CTGGAGCGCATGCGCTAT 5923

   207 AspPheArgIleHisGlyThrIleGluCysTrpLysThrProAlaGlnPh 223
       :::           |||:::::::::                     ||
  5924 GAG............GGCGTGGTCGAC.................ATGTT 5943

   223 eValGlnLeuLysAlaMetGlnAlaGln 232
       |      :::|||:::::::::::::|||
  5944 TCAGACCGTGAAGACCCTGCGTACACAG 5971
```

seq_name: gb_ov:XLU15287

seq_documentation_block:

```
LOCUS       XLU15287     2416 bp    mRNA            VRT       04-MAY-1995
DEFINITION  Xenopus laevis phosphotyrosyl-protein phosphatase mRNA, complete
            cds.
ACCESSION   U15287
NID         g601781
VERSION     U15287.1  GI:601781
KEYWORDS    .
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
            Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae;
            Xenopus.
REFERENCE   1  (bases 1 to 2416)
  AUTHORS   Tang,T.L., Freeman,R.M. Jr., O'Reilly,A.M., Neel,B.G. and
            Sokol,S.Y.
  TITLE     The SH2-containing protein-tyrosine phosphatase SH-PTP2 is required
            upstream of MAP kinase for early Xenopus development
  JOURNAL   Cell 80 (3), 473-483 (1995)
  MEDLINE   95163101
REFERENCE   2  (bases 1 to 2416)
  AUTHORS   Freeman,R.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-SEP-1994) Robert M. Freeman, Molecular Medicine, Beth
            Israel Hospital, 330 Brookline Ave., Boston, MA 02215, USA
FEATURES             Location/Qualifiers
     source          1. .2416
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /clone="pBS-X13a1"
                     /tissue_type="oocyte"
                     /dev_stage="immature egg"
     5'UTR           <1. .113
     CDS             114. .1901
                     /EC_number="3.1.3.48"
                     /note="SH-PTP2"
                     /codon_start=1
                     /product="phosphotyrosyl-protein phosphatase"
                     /protein_id="AAA65731.1"
                     /db_xref="PID:g601782"
                     /db_xref="GI:601782"
                     /translation="MTSRRWFHPNITGVEAENLLLTRGVDGSFLARPSKSNPGDFTLS
                     VRRNGAVTHIKIQNTGDYYDLYGGEKFATLAELVQYYMEHHGQLKEKNGDVIELKYPL
                     NCADPTSERWFHGHLSGKEAEKLLTEKGKHGSFLVRESQSHPGDFVLSVRTGDDKTES
                     NDGKSKVTHVMIRCQDIKYDVGGGEKFDSLTDLVEHYKKNPMVETLGTVLQLKQPLNT
                     TRINAAEIEIRVKELSKPAETADKFKQGFWEEFETLQQQECKLLYSRKEGQRPENKNK
                     NRYKNILPFDHTRVELHDGDPNEQVSDYINANIIMPEFEIKCINPKPKKRYIATQGCL
                     QNTVNDFWRMVFQENSRVIVMTTKEVERGKSKCVKYWPDEFALKEYGVMRVRNVKETP
                     AHDYILRELKLSKVGQGNTERAVWQYHFKTWPDHGVPADPGGVLDFLEEVHHKQDGIT
                     DAGPVVVHCSAGIGRTGTFIVIDILIDVIREKGVDCDIDVPKTIQMVRSQRSGMVQTE
                     AQYRFIYMAVQHYIETLQRRIEEEQKSKRKGHEYTNIKYSLSEQGDQSPLPPCTPTPL
                     PTPTLPEMREDAGRVYENVGLLQQQKSFR"
     misc_feature    135. .434
                     /note="encodes SH2 domain"
     misc_feature    453. .758
                     /note="encodes SH2 domain"
     misc_feature    921. .1693
                     /note="encodes phosphatase domain"
     3'UTR           1902. .>2416
BASE COUNT      801 a    448 c    569 g    598 t
ORIGIN
```

```
alignment_scores:
            Quality:   89.50              Length:    250
              Ratio:  0.772                Gaps:     10
 Percent Similarity:  46.400   Percent Identity: 20.000

alignment_block:
 US-09-068-804-12 x XLU15287   ..

 Align seg 1/1  to: XLU15287  from: 1  to: 2416

     14 GlnGlnThrGlnValLysLeuSerAspGlyMetPro..... 25
        :::::::|||:::|||   |||   ||||||   |||
    969 GATCACACAAGGGTCGAGCTTCATGATGGGGACCCCAATGAACAGGTCTC 1018

     26 .............ValProValIleArgLeuThrPheAspGlyLysProV 38
                     :::   :::|||         |||:::   |||   :
   1019 TGATTACATCAATGCCAACATTATTATGCCTGAATTTGAAATAAAATGCA 1068

     38 alAlaLeuAlaGly............IleArgThrGlnLysIleArgPro 50
        :::::                     |||   ||||||          :::
   1069 TCAATCCCAAACCCAAGAAACGCTACATTGCTACACAAGGCTGCCTACAA 1118

     51 AspArgLeuGluAlaHisMetLysMetLeuLeuGluLysGluCysSerCy 67
        :::   :::::::::      :::|||:::::::::   ::::::
   1119 AACACAGTGAATGATTTCTGGAGAATGGTGTTTCAAGAAAATTCCAGAGT 1168

     67 sLeuValValLeuThrSerGluArgSerAspAlaGlyLysThrIleThrT 84
        :::|||:::   |||   |||         :::      :::   ::::
   1169 TATTGTCATGACCACAAAGGAAGTTGAAAGAGGAAAGAGTAAATGTGTGA 1218

     84 hrTyrPheArgGlySerTyrThr.........PheGlyGluValHisThr 97
        ::|||      :::   ::::::         :::|||   :::::::::
   1219 AGTACTGGCCAGATGAGTTTGCTTTAAAGGAATATGGGGTTATGCGTGTA 1268

     98 AsnSerGlnLysValSerSerAla.......................... 105
           :::   |||   ::::::|||
   1269 AGGAATGTTAAAGAAACTCCAGCTCATGACTATATACTAAGGGAGCTTAA 1318

    106 .............SerGlnGlyGlu.........AlaIleAspGlnTyrA 115
                     :::||||||:::         |||:::   ||||||
   1319 GCTTTCCAAAGTTGGGCAAGGGAACACAGAAAGGGCAGTATGGCAGTAC. 1367

    115 snMetGlnLeuSerCysGlyGluLysArgTyrThrIleProValLeuHis 131
                                                        |||
   1368 ..............................................CAT 1370

    132 ValLysAsnTrpProAspHis...GlnProLeuProSerThrAspGlnLe 147
```

```
1371 TTTAAAACATGGCCAGATCATGGTGTTCCAGCTGACCCTGGTGGAGTCTT 1420

 147 uGluTyrLeuAlaAspArgValLysAsnSerAsnGlnAsnGlyAlaProG 164
1421 GGACTTTTAGAGGAA......GTACACCACAAACAAGATGGTATCACAG 1464

 164 lyArgSerSerSerAspLysHisLeuProMetIleHisCysLeuGlyGly 180
1465 ATGCTGGGCCTGTT..............GTTGTTCATTGCAGTGCTGGA 1499

 181 ValGlyArgThrGlyThrMetAlaAlaAlaLeuValLeuLysAsp..... 195
1500 ATTGGACGGACAGGAACCTTTATTGTGATTGATATTCTTATTGATGTAAT 1549

 195 ..................................................... 195

1550 CAGAGAGAAAGGTGTTGACTGTGACATAGATGTCCCCAAAACTATTCAGA 1599

 196 ........AsnProHisSerAsnLeuGluGlnValArgAlaAspPheArg 209
1600 TGGTGAGATCTCAGAGGTCAGGAATGGTTCAAACAGAAGCACAATACAGA 1649
```

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 1999, 17:14:01 ; Search time 1495.71 Seconds
(without alignments)
763.582 Million cell updates/sec

Title: US-09-068-804-10
Perfect score: 579
Sequence: 1 ATCAGCTTGCCGTCGTCATA.....CCAGCGCATCCCGGCGATTT 579

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 986266752 residues

Database : EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	35.4	6.1	428	26	W47309	W47309 zc40f01.r1
2	34.8	6.0	430	23	R99705	R99705 yq72f05.r1
c 3	34.8	6.0	483	25	N46824	N46824 yy52h06.r1
4	34	5.9	364	42	AI152803	AI152803 ud91c10.r
c 5	33.6	5.8	470	21	T78814	T78814 yc96e03.r1
c 6	33.6	5.8	609	22	R54196	R54196 yg98e10.r1
c 7	33.6	5.8	505	22	R56427	R56427 yg90e12.r1
8	33.2	5.7	347	24	H69399	H69399 yr87e03.r1
9	33	5.7	556	49	AI638165	AI638165 ts97b09.x
c 10	32.6	5.6	334	21	F13334	F13334 HSC3LH031 n
c 11	31.6	5.5	313	29	AA157245	AA157245 zo42b01.s
c 12	31.4	5.4	795	42	AI119592	AI119592 uf05c06.y
13	31	5.4	434	20	D19609	D19609 HUMGS00558
c 14	31	5.4	458	28	AA114966	AA114966 zk89f08.s
c 15	30.6	5.3	372	22	R62330	R62330 yi17f10.s1
16	30.4	5.3	198	21	T79001	T79001 yd21h06.r1
c 17	30.4	5.3	501	39	C88159	C88159 C88159 Mous
c 18	30.4	5.3	422	46	AI442006	AI442006 sa82h11.y
19	30.2	5.2	498	26	W60247	W60247 zd29h01.r1
20	30.2	5.2	448	26	W90434	W90434 zh75f02.r1
21	30.2	5.2	368	28	C15894	C15894 C15894 Clon
22	30.2	5.2	273	29	AA127773	AA127773 zk92a08.r
23	30.2	5.2	222	29	AA148586	AA148586 zl25h10.r
24	30.2	5.2	426	37	AA679503	AA679503 ac55a09.s
c 25	30.2	5.2	467	39	AJ227683	AJ227683 CIC60F Ci
26	30	5.2	469	24	H68166	H68166 yu55b09.r1
27	30	5.2	454	46	AI449079	AI449079 ms91a04.x
28	29.8	5.1	281	50	AV036964	AV036964 AV036964
29	29.6	5.1	242	35	AA586359	AA586359 nd45c09.s
30	29.4	5.1	398	21	D46597	D46597 RICS11364A
31	29.4	5.1	400	31	AA278927	AA278927 zs78c07.r
c 32	29.4	5.1	577	34	AA521404	AA521404 aa68c03.s
33	29.4	5.1	782	42	AI073396	AI073396 ov46a07.x
34	29.4	5.1	469	44	AI274307	AI274307 ql46e10.x
c 35	29.4	5.1	512	51	AI723311	AI723311 fc35e07.x
36	29	5.0	352	27	AA050069	AA050069 mj40h01.r
c 37	29	5.0	343	28	AA093686	AA093686 cl0571.se
38	29	5.0	478	30	AA207547	AA207547 mv79d07.r
39	29	5.0	592	30	AA218241	AA218241 mv61f07.r
40	29	5.0	481	30	AA237737	AA237737 mx29g02.r
41	29	5.0	520	35	AA568788	AA568788 nm07a11.s
42	29	5.0	436	38	AA755907	AA755907 vu04g09.r
43	29	5.0	603	39	AA843510	AA843510 ak08a01.s
44	29	5.0	793	48	AI559540	AI559540 tq50c07.x
c 45	28.8	5.0	381	26	W23722	W23722 zc71a11.s1

ALIGNMENTS

RESULT 1
W47309
LOCUS W47309 428 bp mRNA EST 11-OCT-1996
DEFINITION zc40f01.r1 Soares_senescent_fibroblasts_NbHSF Homo sapiens cDNA clone IMAGE:324793 5', mRNA sequence.
ACCESSION W47309
NID g1331948
VERSION W47309.1 GI:1331948

```
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 428)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     On May 18, 1995 this sequence version replaced.

            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 506   Std Error: 0.00
            Seq primer: mob.REGA+ET
            High quality sequence stop: 374.
FEATURES             Location/Qualifiers
     source          1. .428
                     /organism="Homo sapiens"
                     /db_xref="GDB:1256305"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:324793"
                     /clone_lib="Soares_senescent_fibroblasts_NbHSF"
                     /tissue_type="senescent fibroblast"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Vector: pT7T3D (Pharmacia) with a modified
                     polylinker V_TYPE: phagemid; Site_1: Not I; Site_2: Eco
                     RI; 1st strand cDNA was primed with a Not I - oligo(dT)
                     primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT       99 a     71 c     93 g    165 t
ORIGIN
```

```
  Query Match             6.1%;  Score 35.4;  DB 26;  Length 428;
  Best Local Similarity   55.2%;  Pred. No. 0.66;
  Matches   69;  Conservative    0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy     316 GCCAATTTTTCACATGCAATACCGGGATTGTATACCGCTTTCCCCGCAGGACAGTTGCAT 375
           | || || | || |||  |||   ||| || |  |  || ||||| |    |  |||| |
Db     257 GGCATTTATGCATATGTTATATGCGGACTGCACCCACCTCTCCCCCCCAAGCCTTTGCCT 316

Qy     376 ATTGTATTGGTCTATCGCTTCTCCCTGACTGGCTGAGCTCTCTTTTTGGCTGTTGGTATG 435
           |||  ||| | |   ||||  |||| |||     |  | ||| |      |||||  |
Db     317 CTTGGGTTGTTGTGCTGCTTTCCCCTTACTTTGCTACATTTCTATAGTTAAGTTGGGTTT 376

Qy     436 CACCT 440
           ||| |
Db     377 TACTT 381
```

```
RESULT   2
R99705
LOCUS       R99705         430 bp    mRNA            EST       14-SEP-1995
DEFINITION  yq72f05.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone
            IMAGE:201345 5', mRNA sequence.
ACCESSION   R99705
```

```
NID         g986306
VERSION     R99705.1  GI:986306
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 430)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     On Nov 29, 1993 this sequence version replaced gi:635027.

            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 1589
            High quality sequence stops: 289
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1589   Std Error: 0.00
            Seq primer: M13RP1
            High quality sequence stop: 289.
FEATURES             Location/Qualifiers
     source          1. .430
                     /organism="Homo sapiens"
                     /db_xref="GDB:3770395"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:201345"
                     /clone_lib="Soares fetal liver spleen 1NFLS"
                     /sex="male"
                     /dev_stage="20 week-post conception fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                     with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                     1st strand cDNA was primed with a Pac I - oligo(dT) primer
                     [5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Pac I and cloned into the Pac I
                     and Eco RI sites of the modified pT7T3 vector.  Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      111 a     94 c    112 g    103 t     10 others
ORIGIN
```

```
  Query Match             6.0%;  Score 34.8;  DB 23;  Length 430;
  Best Local Similarity   58.3%;  Pred. No. 1;
  Matches   60;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy     444 CAAAGGTGTAGCTCCCTCTGAAATAGGTGGTAATTGTTTTGCCTGCATCTGATCTTCCGA 503
           ||||     | || ||| |||    ||  ||||| ||| | |||||  |    |  | |
Db      96 CAAAATCACATCTTCCTGTGATCCGGGCTGTAATAGTTCTTCCTGCTTTCTGTNGTGCAG 155

Qy     504 CGTTAACACCACCAGGCACGAGCATTCTTTTTCAAGAAGCATT 546
           | | |||||  |  ||||   ||||  || | | | |||| |
Db     156 AGGTTACACCTCTTGGCACATACATTTCTTCTGATGCAGCAGT 198
```

```
RESULT   3
N46824/c
LOCUS       N46824         483 bp    mRNA            EST       14-FEB-1996
DEFINITION  yy52h06.r1 Soares_multiple_sclerosis_2NbHMSP Homo sapiens cDNA
            clone IMAGE:277211 5', mRNA sequence.
ACCESSION   N46824
```

```
NID         g1187990
VERSION     N46824.1  GI:1187990
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 483)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     On May 5, 1995 this sequence version replaced gi:797732.

            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: T7
            High quality sequence stop: 209.
FEATURES             Location/Qualifiers
     source          1. .483
                     /organism="Homo sapiens"
                     /db_xref="GDB:3895587"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:277211"
                     /clone_lib="Soares_multiple_sclerosis_2NbHMSP"
                     /sex="male"
                     /tissue_type="multiple sclerosis lesions"
                     /dev_stage="Age 46"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Vector: pT7T3D (Pharmacia) with a modified
                     polylinker V_TYPE: phagemid; Site_1: Not I; Site_2: Eco
                     RI; 1st strand cDNA was primed with a Not I - oligo(dT)
                     primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by Bento
                     Soares and M.Fatima Bonaldo. RNA from 4 multiple sclerosis
                     lesions from one patient was kindly provided by Dr. Kevin
                     G. Becker (NINDS/NIH). "
BASE COUNT      109 a    131 c    121 g    112 t     10 others
ORIGIN
```

```
  Query Match             6.0%;  Score 34.8;  DB 25;  Length 483;
  Best Local Similarity   47.5%;  Pred. No. 1.1;
  Matches   96;  Conservative    0;  Mismatches  106;  Indels    0;  Gaps    0;

Qy     339 GGGATTGTATACCGCTTTCCCCGCAGGACAGTTGCATATTGTATTGGTCTATCGCTTCTC 398
           || ||     ||| ||  |||    |   |  || |     |   |||  || ||  ||||
Db     407 GGTATGTAGCACCCCTCCCCCNAAATTCCCCTTCCTCTATTCTTTGCCCTCTCATTTCTG 348

Qy     399 CCTGACTGGCTGAGCTCTCTTTTTGGCTGTTGGTATGCACCTCGCCAAAGGTGTAGCTCC 458
           |   | | ||| |||  ||| | || |  ||||  | |    |  |||   | | |
Db     347 CAGAAGGGNCTGGGCTAGCTTCTGGGATTCTGGTTGGGAAACAGGAAAAACAGAAACCAA 288

Qy     459 CTCTGAAATAGGTGGTAATTGTTTTGCCTGCATCTGATCTTCCGACGTTAACACCACCAG 518
           | |  ||    | |  || | ||    || || ||  ||| | | | ||   | |  |
Db     287 GTTTNCAAGCTCTTGGTATAGGTTNCAGTGGATGTGGGCTTGCTAGGGTATGGCGAGAAC 228

Qy     519 GCACGAGCATTCTTTTTCAAGA 540
               | ||    || |||||
```

```
Db     227 AGGTTGGGATCGGATTGCAAGA 206
```

```
RESULT   4
AI152803
LOCUS       AI152803      364 bp    mRNA            EST       30-SEP-1998
DEFINITION  ud91c10.r1 Soares mouse uterus NMPu Mus musculus cDNA clone
            IMAGE:1478226 5', mRNA sequence.
ACCESSION   AI152803
NID         g3681272
VERSION     AI152803.1  GI:3681272
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 364)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1400880.

            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:926582
            Seq primer: -28m13 rev2 ET from Amersham.
FEATURES             Location/Qualifiers
     source          1. .364
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1478226"
                     /clone_lib="Soares mouse uterus NMPu"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: uterus; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; 1st strand cDNA was prepared from
                     pregnant mouse uterus, and was then primed with a Not I -
                     oligo(dT) primer. Double-stranded cDNA was ligated to Eco
                     RI adaptors (Pharmacia), digested with Not I and cloned
                     into the Not I and Eco RI sites of the modified pT7T3
                     vector.  Library is normalized. Library was constructed by
                     Bento Soares and M. Fatima Bonaldo. "
BASE COUNT       94 a     88 c     79 g    103 t
ORIGIN
```

```
  Query Match             5.9%;  Score 34;  DB 42;  Length 364;
  Best Local Similarity   61.1%;  Pred. No. 1.7;
  Matches   55;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy     455 CTCCCTCTGAAATAGGTGGTAATTGTTTTGCCTGCATCTGATCTTCCGACGTTAACACCA 514
           ||| ||||| | |   |  || | |  || | ||| | ||||  || |   || |||
Db      55 CTCTCTCTGTATTCTTTTATACTGGGGTTTCATGCTTTTGATTCTCTGGGATTTACATTT 114

Qy     515 CCAGGCACGAGCATTCTTTTTCAAGAAGCA 544
           || ||   || || ||||| ||||| | ||
Db     115 CCTGGTGGGACCAATCTTTCTCAAGTACCA 144
```

```
RESULT   5
T78814/c
```

```
LOCUS       T78814         470 bp    mRNA            EST       07-MAR-1995
DEFINITION  yc96e03.r1 Soares infant brain 1NIB Homo sapiens cDNA clone
            IMAGE:23641 5', mRNA sequence.
ACCESSION   T78814
NID         g697323
VERSION     T78814.1  GI:697323
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 470)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 1397
            High quality sequence stops: 420 Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1397   Std Error: 0.00
            Seq primer: M13RP1
            High quality sequence stop: 420.
FEATURES             Location/Qualifiers
     source          1. .470
                     /organism="Homo sapiens"
                     /db_xref="GDB:395988"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:23641"
                     /clone_lib="Soares infant brain 1NIB"
                     /sex="female"
                     /dev_stage="73 days post natal"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: whole brain; Vector: Lafmid BA; Site_1: Not
                     I; Site_2: Hind III; 1st strand cDNA was primed with a Not
                     I - oligo(dT) primer [5'
                     AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTT 3']
                     double-stranded cDNA was ligated to Hind III adaptors
                     (Pharmacia), digested with Not I and directionally cloned
                     into the Not I and Hind III sites of the Lafmid BA vector.
                     Library went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      102 a    118 c    124 g    120 t      6 others
ORIGIN
```

```
  Query Match             5.8%;  Score 33.6;  DB 21;  Length 470;
  Best Local Similarity   51.3%;  Pred. No. 2.5;
  Matches   78;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy     314 TGGCCAATTTTTCACATGCAATACCGGGATTGTATACCGCTTTCCCCGCAGGACAGTTGC 373
           | |||| |||| |   ||| |  || | |  |   ||| |||| |||    | |  |  |
Db     213 TTGCCAGTTTTCCCAGTGCTAGTCCTGCACAGCTCACCTCTTTTCCCTGGCGTCTCTGAC 154

Qy     374 ATATTGTATTGGTCTATCGCTTCTCCCTGACTGGCTGAGCTCTCTTTTTGGCTGTTGGTA 433
             |  |   |||     |      |   |||  |  || ||  || ||   || | ||  ||
Db     153 CTTGTTCCTTGGGACCTTTGGACCTGCTGGTTAACTTAGGCCTTTTGAGGGTTCTTAATA 94

Qy     434 TGCACCTCGCCAAAGGTGTAGCTCCCTCTGAA 465
           |     | |   ||| ||||  ||| |||||
Db      93 AGACAAGCTCACGAGGGGTAGGGCCCCCTGAA 62
```

```
RESULT   6
R54196/c
LOCUS       R54196         609 bp    mRNA            EST       18-MAY-1995
DEFINITION  yg98e10.r1 Soares infant brain 1NIB Homo sapiens cDNA clone
            IMAGE:41515 5', mRNA sequence.
ACCESSION   R54196
NID         g816098
VERSION     R54196.1  GI:816098
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 609)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 414
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: M13RP1
            High quality sequence stop: 414.
FEATURES             Location/Qualifiers
     source          1. .609
                     /organism="Homo sapiens"
                     /db_xref="GDB:414056"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:41515"
                     /clone_lib="Soares infant brain 1NIB"
                     /sex="female"
                     /dev_stage="73 days post natal"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: whole brain; Vector: Lafmid BA; Site_1: Not
                     I; Site_2: Hind III; 1st strand cDNA was primed with a Not
                     I - oligo(dT) primer [5'
                     AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTT 3']
                     double-stranded cDNA was ligated to Hind III adaptors
                     (Pharmacia), digested with Not I and directionally cloned
                     into the Not I and Hind III sites of the Lafmid BA vector.
                     Library went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      130 a    153 c    164 g    152 t      10 others
ORIGIN
```

```
  Query Match             5.8%;  Score 33.6;  DB 22;  Length 609;
  Best Local Similarity   51.3%;  Pred. No. 2.8;
  Matches   78;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy     314 TGGCCAATTTTTCACATGCAATACCGGGATTGTATACCGCTTTCCCCGCAGGACAGTTGC 373
           | |||| |||| |   ||| |  || | |  |   ||| |||| |||    | |  |  |
Db     220 TTGCCAGTTTTCCCAGTGCTAGTCCTGCACAGCTCACCTCTTTTCCCTGGCGTCTCTGAC 161

Qy     374 ATATTGTATTGGTCTATCGCTTCTCCCTGACTGGCTGAGCTCTCTTTTTGGCTGTTGGTA 433
             |  |   ||||     |      |   |||  |  || ||  || ||   || | ||  ||
Db     160 CTTGTTCCTTGGGACCTTTGGACCTGCTGGTTAACTTAGGCCTTTTGAGGGTTCTTAATA 101

Qy     434 TGCACCTCGCCAAAGGTGTAGCTCCCTCTGAA 465
           |     | |   ||| ||||  ||| |||||
Db     100 AGACAAGCTCACGAGGGGTAGGGCCCCCTGAA 69
```

```
RESULT   7
R56427/c
LOCUS       R56427         505 bp    mRNA            EST       23-MAY-1995
DEFINITION  yg90e12.r1 Soares infant brain 1NIB Homo sapiens cDNA clone
            IMAGE:40663 5', mRNA sequence.
ACCESSION   R56427
NID         g826533
VERSION     R56427.1  GI:826533
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 505)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     On Apr 5, 1995 this sequence version replaced gi:760771.

            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 1364
            High quality sequence stops: 420 Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1364   Std Error: 0.00
            Seq primer: M13RP1
            High quality sequence stop: 420.
FEATURES             Location/Qualifiers
     source          1. .505
                     /organism="Homo sapiens"
                     /db_xref="GDB:413204"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:40663"
                     /clone_lib="Soares infant brain 1NIB"
                     /sex="female"
                     /dev_stage="73 days post natal"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: whole brain; Vector: Lafmid BA; Site_1: Not
                     I; Site_2: Hind III; 1st strand cDNA was primed with a Not
                     I - oligo(dT) primer [5'
                     AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTT 3']
                     double-stranded cDNA was ligated to Hind III adaptors
                     (Pharmacia), digested with Not I and directionally cloned
                     into the Not I and Hind III sites of the Lafmid BA vector.
                     Library went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      111 a    133 c    137 g    123 t      1 others
ORIGIN


  Query Match             5.8%;  Score 33.6;  DB 22;  Length 505;
  Best Local Similarity   51.3%;  Pred. No. 2.6;
  Matches   78;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy     314 TGGCCAATTTTTCACATGCAATACCGGGATTGTATACCGCTTTCCCCGCAGGACAGTTGC 373
           | |||| |||| |   ||| |  || | |   |   ||| |||| |||    | |  |  |
Db     227 TTGCCAGTTTTCCCAGTGCTAGTCCTGCACAGCTCACCTCTTTTCCCTGGCGTCTCTGAC 168

Qy     374 ATATTGTATTGGTCTATCGCTTCTCCCTGACTGGCTGAGCTCTCTTTTTGGCTGTTGGTA 433
           |  |   |||;    |     |   |||  |  || ||  || ||   || | ||  ||
Db     167 CTTGTTCCTTGGGACCTTTGGACCTGCTGGTTAACTTAGGCCTTTTGAGGGTTCTTAATA 108

Qy     434 TGCACCTCGCCAAAGGTGTAGCTCCCTCTGAA 465
```

```
           |     | |   ||| |||| |||  |||||
Db     107 AGACAAGCTCACGAGGGGTAGGGCCCCCTGAA 76


RESULT   8
H69399
LOCUS       H69399         347 bp    mRNA            EST       24-OCT-1995
DEFINITION  yr87e03.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone
            IMAGE:212284 5', mRNA sequence.
ACCESSION   H69399
NID         g1039605
VERSION     H69399.1  GI:1039605
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 347)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,
            Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
            Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
            Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J.,
            Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.
            and Marra,M.
  TITLE     Generation and analysis of 280,000 human expressed sequence tags
  JOURNAL   Genome Res. 6 (9), 807-828 (1996)
  MEDLINE   97044478
COMMENT     On Apr 14, 1993 this sequence version replaced gi:716727.

            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 1576
            High quality sequence stops: 235
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1576   Std Error: 0.00
            Seq primer: M13RP1
            High quality sequence stop: 235.
FEATURES             Location/Qualifiers
     source          1. .347
                     /organism="Homo sapiens"
                     /db_xref="GDB:3785125"
                     /db_xref="taxon:9606"
                     /map="954E04; 1; 1q12-1q21.3"
                     /clone="IMAGE:212284"
                     /clone_lib="Soares fetal liver spleen 1NFLS"
                     /sex="male"
                     /dev_stage="20 week-post conception fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                     with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                     1st strand cDNA was primed with a Pac I - oligo(dT) primer
                     [5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Pac I and cloned into the Pac I
                     and Eco RI sites of the modified pT7T3 vector.  Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT       87 a     74 c     92 g     88 t      6 others
ORIGIN


  Query Match             5.7%;  Score 33.2;  DB 24;  Length 347;
  Best Local Similarity   57.8%;  Pred. No. 3.1;
  Matches   59;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy     444 CAAAGGTGTAGCTCCCTCTGAAATAGGTGGTAATTGTTTTGCCTGCATCTGATCTTCCGA 503
```

```
Db     87 CAAAATCACATCTTCCTGTGATCCGGGCTGTAATAGTTCTTCCTGCTTTCTGTTGTGCAG 146

Qy    504 CGTTAACACCACCAGGCACGAGCATTCTTTTTCAAGAAGCAT 545
Db    147 AGGTTACAACTCTTGGCACATACATTTCTTCTGATGCAGCAT 188


RESULT   9
AI638165
LOCUS       AI638165      556 bp    mRNA            EST       27-APR-1999
DEFINITION  ts97b09.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2239193 3',
            mRNA sequence.
ACCESSION   AI638165
NID         g4690399
VERSION     AI638165.1  GI:4690399
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 556)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Jun 22, 1998 this sequence version replaced gi:3247333.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            R. Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco
            High quality sequence stop: 457.
FEATURES             Location/Qualifiers
     source          1. .556
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2239193"
                     /clone_lib="NCI_CGAP_GC6"
                     /tissue_type="pooled germ cell tumors"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Plasmid DNA from the normalized library
                     NCI_CGAP_GC4 was prepared, and ss circles were made in
                     vitro. Following HAP purification, this DNA was used as
                     tracer in a subtractive hybridization reaction. The driver
                     was PCR-amplified cDNAs from a pool of 5,000 clones made
                     from the same library (cloneIDs 1257096-1258631,
                     1469064-1470983, and 1475592-1476743). Subtraction by
                     Bento Soares and M. Fatima Bonaldo.  "
BASE COUNT      139 a    112 c    156 g    149 t
ORIGIN


  Query Match             5.7%;  Score 33;  DB 49;  Length 556;
  Best Local Similarity   54.5%;  Pred. No. 4.1;
  Matches   66;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy    444 CAAAGGTGTAGCTCCCTCTGAAATAGGTGGTAATTGTTTTGCCTGCATCTGATCTTCCGA 503
Db    244 CAAAATCACATCTTCCTGTGATCCGGGCTGTAATAGTTCTTCCTGCTTTCTGTTGTGCAG 303

Qy    504 CGTTAACACCACCAGGCACGAGCATTCTTTTTCAAGAAGCATTTTCATATGCGCTTCCAG 563
```

```
Db    304 AGGTTACAACTCTTGGCACATACATTTCTTCTGATGCAGCAGTAGCTGCAGATCCTGCAC 363

Qy    564 C 564
Db    364 C 364


RESULT  10
F13334/c
LOCUS       F13334        334 bp    mRNA            EST       15-MAR-1995
DEFINITION  HSC3LH031 normalized infant brain cDNA Homo sapiens cDNA clone
            c-3lh03, mRNA sequence.
ACCESSION   F13334
NID         g710207
VERSION     F13334.1  GI:710207
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 334)
  AUTHORS   Auffray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C.,
            Devignes,M.D., Duprat,S., Houlgatte,R., Jumeau,M.N., Lamy,B.,
            Lorenzo,F., Mitchell,H., Mariage-Samson,R., Pietu,G., Pouliot,Y.,
            Sebastiani-Kabaktchis,C. and Tessier,A.
  TITLE     IMAGE: molecular integration of the analysis of the human genome
            and its expression
  JOURNAL   C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
  MEDLINE   95277534
COMMENT
            Contact: Genethon
            Genexpress-Genethon
            Genethon Centre de recherche sur le Genome Humain
            1,rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
            Tel: 33169472800
            Fax: 33160778698
            Email: genexpress@genethon.fr
            Single read.
            Genexpress_library_idt: C; Genexpress_sequence_idt: y1c-3lh03
            Insert Length: 1229   Std Error: 0.00
            Seq primer: (-21)M13_universal
            High quality sequence stop: 265.
FEATURES             Location/Qualifiers
     source          1. .334
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="22"
                     /clone="c-3lh03"
                     /clone_lib="normalized infant brain cDNA"
                     /sex="Female"
                     /tissue_type="total brain"
                     /dev_stage="3 months old"
                     /note="Organ: brain; Vector: lafmid BA; Site_1: HindIII;
                     Site_2: NotI; sex=Female; dev_stage=3 months old;
                     isolate=muscular atrophy patient;  tissue_type=total
                     brain;  total mRNA was oligo-(dT) primed and directionally
                     cloned 5' -> 3' into the  HindIII -> NotI sites of the
                     lafmid BA vector. Clone library from B.Soares,  Psychiatry
                     Dept. Columbia University, USA.  Normalization_method:
                     Bento Soares, P.N.A.S in press"
BASE COUNT       75 a     90 c     91 g     76 t      2 others
ORIGIN


  Query Match             5.6%;  Score 32.6;  DB 21;  Length 334;
  Best Local Similarity   50.7%;  Pred. No. 4.7;
  Matches   77;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy    314 TGGCCAATTTTTCACATGCAATACCGGGATTGTATACCGCTTTCCCCCGCAGGACAGTTGC 373
Db    228 TTGCCAGTTTTCCCAGTGCTAGTCCTGCACAGCTCACCTCTTTTCCCTGGCGTCTCTGAC 169
```

```
Qy     374 ATATTGTATTGGTCTATCGCTTCTCCCTGACTGGCTGAGCTCTCTTTTTGGCTGTTGGTA 433
Db     168 CTTGTTCCTTGGGACCTTTGGACCTGCTGGTTAACTTAGGCCTTTTGAGGGTTCTTAATA 109

Qy     434 TGCACCTCGCCAAAGGTGTAGCTCCCTCTGAA 465
Db     108 AGACAAGCTCACGAGGGGTANGGCCCCCTGAA 77


RESULT  11
AA157245/c
LOCUS       AA157245      313 bp    mRNA            EST       11-DEC-1996
DEFINITION  zo42b01.s1 Stratagene endothelial cell 937223 Homo sapiens cDNA
            clone IMAGE:589513 3', mRNA sequence.
ACCESSION   AA157245
NID         g1728913
VERSION     AA157245.1  GI:1728913
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 313)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,
            Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
            Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
            Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J.,
            Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.
            and Marra,M.
  TITLE     Generation and analysis of 280,000 human expressed sequence tags
  JOURNAL   Genome Res. 6 (9), 807-828 (1996)
  MEDLINE   97044478
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1393251.

            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40M13 fwd. from Amersham
            High quality sequence stop: 298.
FEATURES             Location/Qualifiers
     source          1. .313
                     /organism="Homo sapiens"
                     /db_xref="GDB:4627253"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:589513"
                     /clone_lib="Stratagene endothelial cell 937223"
                     /dev_stage="umbilical vein, 1 passage"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; Cloned unidirectionally.  Primer: Oligo dT.
                     Umbilical vein endothelial cells, passaged once. Average
                     insert size: 1.0 kb; Uni-ZAP XR Vector; ~5' adaptor
                     sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor sequence: 5'
                     CTCGAGITTTTTTTTTTTTTTTT 3'"
BASE COUNT      130 a     56 c     46 g     81 t
ORIGIN


  Query Match             5.5%;  Score 31.6;  DB 29;  Length 313;
  Best Local Similarity   55.5%;  Pred. No. 9.4;
  Matches   61;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy      28 GGGCTTGCATTGCTTTTAGTTGTACAAACTGTGCAGGCGTCTTCCAGCATTCTATTGTTC 87
Db     224 GGTCTTTCATTTGTCATCGTTTTTGGTGTTTTGCCTAAGTATTTCCATATTTTTATTTTA 165

Qy      88 CGTGAATCCGGAAATCTGCACGTACCTGCTCCAGATTACTATGAGGATTA 137
Db     164 CTTGTCTGTGGAGGTTTGCTATTAAGTTCTGCAGTTTTGTGTGAGGATAA 115


RESULT  12
AI119592/c
LOCUS       AI119592      795 bp    mRNA            EST       02-SEP-1998
DEFINITION  uf05c06.y1 Sugano mouse liver mlia Mus musculus cDNA clone
            IMAGE:1499722 5' similar to gb:X64262 M.musculus mRNA for
            Apolipoprotein A-I (MOUSE);, mRNA sequence.
ACCESSION   AI119592
NID         g3519916
VERSION     AI119592.1  GI:3519916
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 795)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2284655.

            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:937326
            Seq primer: custom primer used
            High quality sequence stop: 526.
FEATURES             Location/Qualifiers
     source          1. .795
                     /organism="Mus musculus"
                     /strain="C57BL"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1499722"
                     /clone_lib="Sugano mouse liver mlia"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: liver; Vector: pME18S-FL3; Site_1: DraIII
                     (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
                     was primed with an oligo(dT) primer
                     [ATGTGGCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was
                     ligated to a  DraIII adaptor [TGTTGGCCTACTGG], digested
                     and cloned into distinct DraIII sites of the pME18S-FL3
                     vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should
                     be used to isolate the cDNA insert.  Size selection was
                     performed to exclude fragments <1.5kb.  Library
                     constructed by Dr. Sumio Sugano (University of Tokyo
                     Institute of Medical Science).  Custom primers  for
                     sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
                     primer CGACCTGCAGCTCGAGCACA."
BASE COUNT      208 a    215 c    239 g    133 t
ORIGIN


  Query Match             5.4%;  Score 31.4;  DB 42;  Length 795;
  Best Local Similarity   61.7%;  Pred. No. 15;
  Matches   50;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy     384 GGTCTATCGCTTCTCCCTGACTGGCTGAGCTCTCTTTTTGGCTGTTGGTATGCACCTCGC 443
```

```
Db    738 GGTCTGGCTTTTCTCGCCAAGTGTCTCAGGTTGGGTTTTGGCCCTTGGTGTGGTACTCGT 679

Qy    444 CAAAGGTGTAGCTCCCTCTGA 464
          | ||||| | | |  |||
Db    678 CCAAGGTAGGGTTGCTCTTGA 658


RESULT 13
D19609
LOCUS       D19609        434 bp    mRNA            EST       30-JUL-1996
DEFINITION  HUMGS00558 Human promyelocyte Homo sapiens cDNA clone mm0472 3',
            mRNA sequence.
ACCESSION   D19609
NID         g499925
VERSION     D19609.1  GI:499925
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 434)
  AUTHORS   Okubo,K., Fukushima,A., Yoshii,J., Niiyama,T., Kojima,Y.,
            Yoshinari,H., Arimoto,J. and Matsubara,K.
  TITLE     Gene expression of human promyelocytic cell line HL60 before and
            after induction of differentiation. A new application of 3'directed
            cDNA sequencing
  JOURNAL   Unpublished (1993)
COMMENT
            Contact: Okubo,K., Fukushima,A., Yoshii,J., Niiyama,T., Kojima,Y.,
            Yoshinari,H., Arimoto,J. and Matsubara,K.
            Institute for Molecular and Cellular Biology
            Osaka University
            3-1 Yamada-oka,Suita,Osaka 565,Japan.
FEATURES             Location/Qualifiers
     source          1. .434
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="mm0472"
                     /clone_lib="Human promyelocyte"
                     /note="Female, adult, cell_line = HL60, cell_type =
                     promyelocyte. "
BASE COUNT      125 a     72 c     79 g    151 t      7 others
ORIGIN


  Query Match             5.4%;  Score 31;  DB 20;  Length 434;
  Best Local Similarity   51.0%;  Pred. No. 16;
  Matches   73;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy    191 ACAATGAATCATCGGTAAATGCTTATCTGATGAACTACGCCCCGGCGCGCCATTTTGGTT 250
          || | ||||||| | | |  | | || ||||  |  |||    |    |||||
Db     83 ACTAAGAATCATTCATGAGTAAATGTGTGTTGAATATCTACCCTTGACCTTTTTTTGAGA 142

Qy    251 ACTATTTTTCACCCTATCCGCCAGGTATTCTAACTGATCCGTAGACGGTAACGGCTGGTG 310
          | ||   |  || | |||     || |   | || ||  | |||  |    | ||  |
Db    143 AATAGAGTAAACACAGTCCCTGTAGTCTGACAGCTAATGGGGAGAGAGGGTAGACTTTTC 202

Qy    311 ATCTGGCCAATTTTTCACATGCA 333
          |||    || |||  |||||||
Db    203 ATCGAATTAAATTTCTACATGCA 225


RESULT 14
AA114966/c
LOCUS       AA114966      458 bp    mRNA            EST       14-MAY-1997
DEFINITION  zk89f08.s1 Soares_pregnant_uterus_NbHPU Homo sapiens cDNA clone
            IMAGE:490023 3', mRNA sequence.
ACCESSION   AA114966
NID         g1670181
VERSION     AA114966.1  GI:1670181
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 458)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,
            Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
            Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
            Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J.,
            Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.
            and Marra,M.
  TITLE     Generation and analysis of 280,000 human expressed sequence tags
  JOURNAL   Genome Res. 6 (9), 807-828 (1996)
  MEDLINE   97044478
COMMENT     On Apr 14, 1993 this sequence version replaced gi:693015.

            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 182   Std Error: 0.00
            Seq primer: -40M13 fwd. from Amersham
            High quality sequence stop: 352.
FEATURES             Location/Qualifiers
     source          1. .458
                     /organism="Homo sapiens"
                     /db_xref="GDB:3804610"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:490023"
                     /clone_lib="Soares_pregnant_uterus_NbHPU"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: uterus; Vector: pT7T3-Pac; Site_1: Not I;
                     Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                     oligo(dT) primer [5'
                     AACTGGAAGAATTCGCGGCCGCCTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization.  Library
                     constructed by M. Fatima Bonaldo."
BASE COUNT      96 a    124 c    138 g     99 t      1 others
ORIGIN


  Query Match             5.4%;  Score 31;  DB 28;  Length 458;
  Best Local Similarity   48.3%;  Pred. No. 17;
  Matches   85;  Conservative    0;  Mismatches   91;  Indels    0;  Gaps    0;

Qy    250 TACTATTTTTCACCCTATCCGCCAGGTATTCTAACTGATCCGTAGACGGTAACGGCTGGT 309
          | |  |||| |     | | ||    |      || |||   ||  | ||| |   ||||
Db    438 TGCAGTTTTACCAGNAAACGGCTCCTTCCCTAAAATGAGAAGTCCAAGGTCATCTCTGGC 379

Qy    310 GATCTGGCCAATTTTTCACATGCAATACCGGGATTGTATACCGCTTTCCCCGCAGGACAG 369
            || |||     |  |||| | |||| ||     | || || || ||  ||
Db    378 CCAGTGACCACAGAGAGATCTGCACTCCCGGACTTCAGGCCTGCCTTTCCAGCGAGAATT 319

Qy    370 TTGCATATTGTATTGGTCTATCGCTTCTCCCTGACTGGCTGAGCTCTCTTTTTGGC 425
          | |||  |  |  | ||  | |||  || |||||   |   ||    |  |||| |
Db    318 CTTCATCCTCCACGGTTCACTAGCTCCTACCTGAAGAGGAAAGGGGGCCATTTGAC 263


RESULT 15
R62330/c
LOCUS       R62330        372 bp    mRNA            EST       26-MAY-1995
DEFINITION  yi17f10.s1 Soares placenta Nb2HP Homo sapiens cDNA clone
            IMAGE:139531 3', mRNA sequence.
ACCESSION   R62330
NID         g834209
```

```
VERSION     R62330.1  GI:834209
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 372)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     On May 9, 1995 this sequence version replaced gi:802856.

            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 705
            High qality sequence stops: 241
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 705   Std Error: 0.00
            Seq primer: Promega -21m13
            High quality sequence stop: 241.
FEATURES             Location/Qualifiers
     source          1. .372
                     /organism="Homo sapiens"
                     /db_xref="GDB:546126"
                     /db_xref="taxon:9606"
                     /map="4"
                     /clone="IMAGE:139531"
                     /clone_lib="Soares placenta Nb2HP"
                     /sex="Female"
                     /dev_stage="placenta obtained at birth (full term)"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: placenta; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo.  "
BASE COUNT      111 a     64 c    117 g     70 t     10 others
ORIGIN


  Query Match             5.3%;  Score 30.6;  DB 22;  Length 372;
  Best Local Similarity   49.3%;  Pred. No. 21;
  Matches   75;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy     254 ATTTTTCACCCTATCCGCCAGGTATTCTAACTGATCCGTAGACGGTAACGGCTGGTGATC 313
           | ||| || || |   | |    | |    |  |||   || || || |  || | ||
Db     340 AGTTTGCAGCCGAGTNGGGACTCCTCCCCTGTCCTCCTACGAAGGAAAAGANTGCTAATG 281

Qy     314 TGGCCAATTTTTCACATGCAATACCGGGATTGTATACCGCTTTCCCCGCAGGACAGTTGC 373
           | |       ||  | | | | | |||  | |     | |||| || |    || ||
Db     280 TCGTACTCACCTCCGACGTACTCCGGGGTCTTTTGGGAGTTTTCTCCCCTAACCANTTCA 221

Qy     374 ATATTGTATTGGTCTATCGCTTCTCCCTGACT 405
           |  || | ||||  | |||||  | | || ||
Db     220 ACTTTTTTTTGGAATCTCGCTCTTGCATGCCT 189


Search completed: October  5, 1999, 17:14:06
```

Job time: 8689 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 1999, 18:01:06 ; Search time 140.41 Seconds
(without alignments)
429.037 Million cell updates/sec

Title: US-09-068-804-10
Perfect score: 579
Sequence: 1 ATCAGCTTGCCGTCGTCATA.....CCAGCGCATCCCGGCGATTT 579

Scoring table: IDENTITY_NUC

Searched: 192659 seqs, 52021692 residues

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/5C_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/5D_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	71.4	12.3	2406	3	US-08-586-272-1	Sequence 1, Appli
	2	48.6	8.4	7218	2	US-08-232-463-14	Sequence 14, Appl
	3	29	5.0	4203	3	US-08-288-630-3	Sequence 3, Appli
c	4	29	5.0	1651	4	US-08-844-058-1	Sequence 1, Appli
c	5	28.6	4.9	2334	1	US-08-406-070-1	Sequence 1, Appli
	6	28.2	4.9	1113	3	US-08-341-538A-1	Sequence 1, Appli
	7	28.2	4.9	1113	3	US-08-725-518-1	Sequence 1, Appli
	8	28	4.8	1591	2	US-08-728-956-3	Sequence 3, Appli
c	9	27.4	4.7	1545	2	US-08-400-275-17	Sequence 17, Appl
c	10	27	4.7	629	3	US-08-670-186-1	Sequence 1, Appli
	11	26.8	4.6	20710	3	US-08-420-235B-1	Sequence 1, Appli
	12	26.8	4.6	984	3	US-08-420-235B-34	Sequence 34, Appl
	13	26.8	4.6	35100	3	US-08-770-379-18	Sequence 18, Appl
	14	26.8	4.6	20710	5	PCT-US95-10194-1	Sequence 1, Appli
	15	26.8	4.6	984	5	PCT-US95-10194-34	Sequence 34, Appl
	16	26.6	4.6	4094	2	US-08-571-758-5	Sequence 5, Appli
	17	26.6	4.6	4094	2	US-08-909-984A-5	Sequence 5, Appli
	18	26.6	4.6	4094	2	US-08-909-983-5	Sequence 5, Appli
c	19	26.6	4.6	2483	3	US-08-177-109A-1	Sequence 1, Appli
c	20	26.6	4.6	2483	4	US-08-687-706-1	Sequence 1, Appli
	21	26.4	4.6	2306	3	US-08-378-698-3	Sequence 3, Appli
c	22	26.4	4.6	3225	3	US-08-482-090-14	Sequence 14, Appl
c	23	26.4	4.6	3225	3	US-08-481-700B-7	Sequence 7, Appli
	24	26.4	4.6	2306	5	PCT-US96-00728-3	Sequence 3, Appli
c	25	26.2	4.5	588	2	US-08-076-091C-6	Sequence 6, Appli
c	26	26.2	4.5	588	2	US-08-285-641-6	Sequence 6, Appli
	27	26	4.5	2713	1	US-08-285-439-1	Sequence 1, Appli
	28	26	4.5	2441	3	US-08-920-812-12	Sequence 12, Appl
	29	26	4.5	2441	3	US-08-920-827-12	Sequence 12, Appl
	30	26	4.5	2441	3	US-08-921-177-12	Sequence 12, Appl
	31	26	4.5	2441	3	US-08-362-577C-12	Sequence 12, Appl
	32	26	4.5	2441	3	US-08-920-828-12	Sequence 12, Appl
c	33	25.8	4.5	1368	1	US-07-677-413-1	Sequence 1, Appli
c	34	25.8	4.5	1571	1	US-07-677-413-2	Sequence 2, Appli
c	35	25.8	4.5	843	3	US-08-040-548-9	Sequence 9, Appli
c	36	25.8	4.5	3086	3	US-08-040-548-15	Sequence 15, Appl
c	37	25.8	4.5	843	3	US-08-466-344-9	Sequence 9, Appli
c	38	25.8	4.5	3086	3	US-08-466-344-15	Sequence 15, Appl
c	39	25.8	4.5	386	3	US-08-620-467A-9	Sequence 9, Appli
c	40	25.8	4.5	386	3	US-08-348-572-9	Sequence 9, Appli
c	41	25.8	4.5	3068	3	US-08-224-482-1	Sequence 1, Appli
	42	25.8	4.5	2658	4	US-08-826-426-14	Sequence 14, Appl
	43	25.8	4.5	513	5	PCT-US92-00282-20	Sequence 20, Appl
c	44	25.6	4.4	1720	3	US-08-457-997B-1	Sequence 1, Appli
	45	25.6	4.4	9385	4	US-08-874-266-1	Sequence 1, Appli

ALIGNMENTS

```
RESULT 1
US-08-586-272-1
; Sequence 1, Application US/08586272
; Patent No. 5824795
;  GENERAL INFORMATION:
;    APPLICANT:  POPOFF, MICHEL Y.
;    APPLICANT:  LE GUERN FELLOUS, MURIEL
;    TITLE OF INVENTION:  OLIGONUCLEOTIDES FOR THE DETECTION OF
;    TITLE OF INVENTION:  SALMONELLA
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
;      CITY:  ARLINGTON
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/586,272
;      FILING DATE:  16-JAN-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 9500410
;      FILING DATE:  16-JAN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OBLON, NORMAN F.
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  660-107-0
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-413-3000
;      TELEFAX:  703-413-2220
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2406 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  SALMONELLA SER. TYPHI
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  97..1755
;      OTHER INFORMATION:  /product= "iagA"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1776..2255
;      OTHER INFORMATION:  /product= "iagB"
US-08-586-272-1


  Query Match            12.3%;  Score 71.4;  DB 3;  Length 2406;
  Best Local Similarity  96.9%;  Pred. No. 1.9e-15;
  Matches   94;  Conservative    0;  Mismatches    1;  Indels    2;  Gaps    2;
```

```
Qy      1 ATCAGCTTGCCGTCGTCATAAGCAACTGG-GCTTGCATTGCTTTTAGTTGTACAAACTGT 59
          |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db   2311 ATCAGCTTGCCGTCGTCATAAGCAACTGGCGCTTGCATTGCTTTTAGTTGTACAAACTGT 2370

Qy     60 GCAGGCGTCTTCCAGCATTCTATTGTTCCGTGAATCC 96
          | ||||||||||||||||||||||||||||||||||| |
Db   2371 G-AGGCGTCTTCCAGCATTCTATTGTTCCGTGAATTC 2406


RESULT   2
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14


  Query Match             8.4%;  Score 48.6;  DB 2;  Length 7218;
  Best Local Similarity   4.5%;  Pred. No. 4e-07;
  Matches   18;  Conservative  216;  Mismatches  165;  Indels    0;  Gaps    0;

Qy    154 CCGCCGCCATCGTTCCGGTTCTTCCCACTCCGCCCAGACAATGAATCATCGGTAAATGCT 213
          :: :: :: :: :::: :::::::: :::: :::   :  :   :: :: :   : ::
Db   1072 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1131

Qy    214 TATCTGATGAACTACGCCCCGGCGCGCCATTTTGGTTACTATTTTTCACCCTATCCGCCA 273
          : :::  :   :: : ::::  : : :: ::::  :: :: :::::: :::: ::: ::
Db   1132 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1191

Qy    274 GGTATTCTAACTGATCCGTAGACGGTAACGGCTGGTGATCTGGCCAATTTTTCACATGCA 333
          : :::: :: ::: :   :  :  :  :: :  :  :: :::: :: :::::: : : :
Db   1192 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1251

Qy    334 ATACCGGGATTGTATACCGCTTTCCCCGCAGGACAGTTGCATATTGTATTGGTCTATCGC 393
          : ::    :: : : :: :::::::: :    :  :: : : :: : ::  ::: :: :
Db   1252 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1311

Qy    394 TTCTCCCTGACTGGCTGAGCTCTCTTTTTGGCTGTTGGTATGCACCTCGCCAAAGGTGTA 453
          :::::::: ::  ::   :::::::::: :: ::  : : :::: ::     : :
Db   1312 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1371

Qy    454 GCTCCCTCTGAAATAGGTGGTAATTGTTTTGCCTGCATCTGATCTTCCGACGTTAACACC 513
          ::::::::    :  :  :  :: :::: ::: : :::  ::::::  : ::  : ::
Db   1372 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1431

Qy    514 ACCAGGCACGAGCATTCTTTTTCAAGAAGCATTTTCATA 552
          ::   |   |   |||| | ||   ||  | || ||||
Db   1432 YYYGTACCAAATTCTTCTATCTCTTTAACTACTTGCATA 1470


RESULT   3
US-08-288-630-3
; Sequence 3, Application US/08288630
; Patent No. 5837848
;  GENERAL INFORMATION:
;    APPLICANT:  ELY, SUSAN
;    APPLICANT:  EVANS, IAN J
;    APPLICANT:  SCHUCH, WOLFGANG W
;    TITLE OF INVENTION:  ROOT-SPECIFIC PROMOTER
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN, L.L.P.
;      STREET:  1100 New York Avenue, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-3918
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/288,630
;      FILING DATE:  10-AUG-1994
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/111,590
;      FILING DATE:  25-AUG-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/669,433
;      FILING DATE:  15-MAR-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9006017.9
;      FILING DATE:  16-MAR-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KOKULIS, PAUL N
;      REGISTRATION NUMBER:  16,773
;      REFERENCE/DOCKET NUMBER:  215964/SEE35669USCIP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 861-3000
;      TELEFAX:  (202) 822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4203 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-288-630-3
```

```
  Query Match             5.0%;  Score 29;  DB 3;  Length 4203;
  Best Local Similarity   55.4%;  Pred. No. 2.5;
  Matches   56;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy     455 CTCCCTCTGAAATAGGTGGTAATTGTTTTGCCTGCATCTGATCTTCCGACGTTAACACCA 514
           ||||   |  || | |   | || |   ||| ||||| |  | |||     || |   |
Db     760 CTCCTATTTTAAAATGCCATCATAGGGGTGCTTGCATTTCTTGTTCATGATTTTATTACT 819

Qy     515 CCAGGCACGAGCATTCTTTTTCAAGAAGCATTTTCATATGC 555
           | || ||  ||  | ||||||  |  | |  || |||||||
Db     820 CAAGTCAAAAGTCTGCTTTTTATATTACCTATTACATATGC 860


RESULT   4
US-08-844-058-1/c
; Sequence 1, Application US/08844058
; Patent No. 5952196
;  GENERAL INFORMATION:
;    APPLICANT:  Lawlor, Elizabeth
;    TITLE OF INVENTION:  No. 5952196el Compounds
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corporation
;      STREET:  709 Swedeland Road
;      CITY:  King of Prussia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19406-0939
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/844,058
;      FILING DATE:  18-APR-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  9607991.8
;      FILING DATE:  18-APR-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gimmi, Edward R
;      REGISTRATION NUMBER:  38,891
;      REFERENCE/DOCKET NUMBER:  P31458-6
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-270-4478
;      TELEFAX:  610-270-5090
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1651 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
US-08-844-058-1


  Query Match             5.0%;  Score 29;  DB 4;  Length 1651;
  Best Local Similarity   57.0%;  Pred. No. 1.5;
  Matches   53;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy     369 GTTGCATATTGTATTGGTCTATCGCTTCTCCCTGACTGGCTGAGCTCTCTTTTTGGCTGT 428
           | |||  ||| | | ||   || | ||| ||  |  |  | |   || | |||||||  |
Db    1562 GGTGCCAATTCTTTAGGGAAATTGATTCCCCAAGCGTAACGGTATTCACCTTTTGGCGTT 1503

Qy     429 TGGTATGCACCTCGCCAAAGGTGTAGCTCCCTC 461
           | ||   ||    | | |  |||| ||||| ||
Db    1502 TTGTTAACAAAGAGGCGAGCGTGTTGCTCCATC 1470


RESULT   5
US-08-406-070-1/c
; Sequence 1, Application US/08406070
; Patent No. 5610063
;  GENERAL INFORMATION:
;    APPLICANT:  Smith, Daniel S.
;    APPLICANT:  Walker, John C.
;    TITLE OF INVENTION:  A cDNA For
;    TITLE OF INVENTION:  alpha-N-acetyl-galactosaminidase From Gallus domesticus
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Reising, Ethington, Barnard, Perry & Milton
;      STREET:  P.O. Box  4390
;      CITY:  Troy
;      STATE:  Michigan
;      COUNTRY:  US
;      ZIP:  48099
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/406,070
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kohn, Kenneth I.
;      REGISTRATION NUMBER:  30,955
;      REFERENCE/DOCKET NUMBER:  UMO P-317
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (810) 689-3500
;      TELEFAX:  (810) 689-4071
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2334 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  2316..2334
US-08-406-070-1


  Query Match             4.9%;  Score 28.6;  DB 1;  Length 2334;
  Best Local Similarity   59.0%;  Pred. No. 2.5;
  Matches   49;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy     446 AAGGTGTAGCTCCCTCTGAAATAGGTGGTAATTGTTTTGCCTGCATCTGATCTTCCGACG 505
           |||||  |||  |    ||   || ||  | |||    ||   |||  |    ||| ||
Db    1863 AAGGTTCAGCCTCACAAGATTGAGATGCAACTTGGCCAGCACCCATGAGCATCTCCCACC 1804

Qy     506 TTAACACCACCAGGCACGAGCAT 528
           |||| |||||||  | ||  |||||
Db    1803 TTAAGACCACCACGAACAAGCAT 1781


RESULT   6
US-08-341-538A-1
; Sequence 1, Application US/08341538A
; Patent No. 5817317
;  GENERAL INFORMATION:
;    APPLICANT:  TITBALL DR, RICHARD W
;    APPLICANT:  WILLIAMSON DR, ETHEL D
;    TITLE OF INVENTION:  CLOSTRIDIUM PERFRINGENS VACCINES
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NIXON & VANDERHYE
;      STREET:  8th FLOOR, 1100 NORTH GLEBE ROAD
;      CITY:  ARLINGTON
;      STATE:  VIRGINIA
;      COUNTRY:  UNITED STATES
;      ZIP:  22201-4714
```

```
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/341,538A
;      FILING DATE:  28-NOV-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9210717.6
;      FILING DATE:  20-MAY-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9215655.3
;      FILING DATE:  23-JUL-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO GB93/01039
;      FILING DATE:  20-MAY-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CRAWFORD, ARTHUR R
;      REGISTRATION NUMBER:  25327
;      REFERENCE/DOCKET NUMBER:  1498-58
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1113 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Clostridium perfringens
;      STRAIN:  double
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1110
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  736..1110
US-08-341-538A-1


  Query Match             4.9%;  Score 28.2;  DB 3;  Length 1113;
  Best Local Similarity   51.2%;  Pred. No. 2.3;
  Matches   66;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy     237 GCGCCATTTTGGTTACTATTTTTCACCCTATCCGCCAGGTATTCTAACTGATCCGTAGAC 296
           || | |||||||  |       | | || ||||  || | || | | ||||       | |
Db     375 GCACTATTTTGGAGATATAGATACTCCATATCATCCTGCTAATGTTACTGCCGTTGATAG 434

Qy     297 GGTAACGGCTGGTGATCTGGCCAATTTTTCACATGCAATACCGGGATTGTATACCGCTTT 356
             | |     || | |  | |  | |||| || | | || |    | |  |||||
Db     435 CGCAGGACATGTTAAGTTTGAGACTTTTGCAGAGGAAAGAAAAGAACAGTATAAAATAAA 494

Qy     357 CCCCGCAGG 365
           | | |||||
Db     495 CACAGCAGG 503


RESULT   7
US-08-725-518-1
; Sequence 1, Application US/08725518
; Patent No. 5851827
;  GENERAL INFORMATION:
;    APPLICANT:  TITBALL DR, RICHARD W
;    APPLICANT:  WILLIAMSON DR, ETHEL D
;    TITLE OF INVENTION:  CLOSTRIDIUM PERFRINGENS VACCINES
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NIXON & VANDERHYE
;      STREET:  8th FLOOR, 1100 NORTH GLEBE ROAD
;      CITY:  ARLINGTON
;      STATE:  VIRGINIA
;      COUNTRY:  UNITED STATES
;      ZIP:  22201-4714
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/725,518
;      FILING DATE:  4-OCT-96
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/341538
;      FILING DATE:  28-NOV-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9210717.6
;      FILING DATE:  20-MAY-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9215655.3
;      FILING DATE:  23-JUL-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO GB93/01039
;      FILING DATE:  20-MAY-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CRAWFORD, ARTHUR R
;      REGISTRATION NUMBER:  25327
;      REFERENCE/DOCKET NUMBER:  124-536
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1113 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Clostridium perfringens
;      STRAIN:  double
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1110
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  736..1110
US-08-725-518-1


  Query Match             4.9%;  Score 28.2;  DB 3;  Length 1113;
  Best Local Similarity   51.2%;  Pred. No. 2.3;
  Matches   66;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy     237 GCGCCATTTTGGTTACTATTTTTCACCCTATCCGCCAGGTATTCTAACTGATCCGTAGAC 296
           || | |||||||  |       | | || ||||  || | || | | ||||       | |
Db     375 GCACTATTTTGGAGATATAGATACTCCATATCATCCTGCTAATGTTACTGCCGTTGATAG 434

Qy     297 GGTAACGGCTGGTGATCTGGCCAATTTTTCACATGCAATACCGGGATTGTATACCGCTTT 356
             | |     || | |  | |  | |||| || | | || |    | |  |||||
Db     435 CGCAGGACATGTTAAGTTTGAGACTTTTGCAGAGGAAAGAAAAGAACAGTATAAAATAAA 494

Qy     357 CCCCGCAGG 365
           | | |||||
Db     495 CACAGCAGG 503


RESULT   8
US-08-728-956-3
; Sequence 3, Application US/08728956
; Patent No. 5677175
;  GENERAL INFORMATION:
;    APPLICANT:  Hodges, Thomas K.
```

```
;    APPLICANT:  McGee, J. D.
;    TITLE OF INVENTION:  Plant Pathogen Induced Proteins
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Barnes & Thornburg
;      STREET:  11 S Meridian
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  USA
;      ZIP:  46204
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/728,956
;      FILING DATE:
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Breen, John P.
;      REGISTRATION NUMBER:  38,833
;      REFERENCE/DOCKET NUMBER:  3220-26119
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (317)231-7745
;      TELEFAX:  (317)231-7433
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1591 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Oryza sativa
US-08-728-956-3


  Query Match             4.8%;  Score 28;  DB 2;  Length 1591;
  Best Local Similarity   56.5%;  Pred. No. 3.3;
  Matches   52;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy     135 TTATCCTTAAGTACAAGGGCCGCCGCCATCGTTCCGGTTCTTCCCACTCCGCCCAGACAA 194
           | |||||| ||   ||| | |  |  |  ||| |   |||| |  || ||    | ||
Db     210 TCATCCTTTAGGCTAAGCTCGGGAGTGAGGGTTGGGAACCTTCACTTTCAGCATGCAAAA 269

Qy     195 TGAATCATCGGTAAATGCTTATCTGATGAACT 226
           |||| |  |||| |  || |  || || || |
Db     270 TGAAGCGGCGGTTAGCGCATGACTAATTAAGT 301


RESULT  9
US-08-400-275-17/c
; Sequence 17, Application US/08400275
; Patent No. 5668295
;  GENERAL INFORMATION:
;    APPLICANT:  Wahab, Samir Z.
;    APPLICANT:  Malik, Vedpal S.
;    TITLE OF INVENTION:  PUTRESCINE N-METHYLTRANSFERASE,
;    TITLE OF INVENTION:  RECOMBINANT DNA MOLECULES ENCODING PUTRESCINE
;    TITLE OF INVENTION:  N-METHYLTRANSFERASE, AND TRANSGENIC TOBACCO PLANTS WITH
;    TITLE OF INVENTION:  ALTERED NICOTINE CONTENT
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Neave
;      STREET:  1251 Ave. of the Americas
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10020
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/400,275
;      FILING DATE:
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/076,681
;      FILING DATE:
;      APPLICATION NUMBER:  US 07/613,160
;      FILING DATE:  14-NOV-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Creason, Gary L
;      REGISTRATION NUMBER:  34,310
;      REFERENCE/DOCKET NUMBER:  PM-1696
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-596-9000
;      TELEFAX:  212-596-9090
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1545 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-400-275-17


  Query Match             4.7%;  Score 27.4;  DB 2;  Length 1545;
  Best Local Similarity   51.2%;  Pred. No. 5.3;
  Matches   64;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy     294 GACGGTAACGGCTGGTGATCTGGCCAATTTTTCACATGCAATACCGGGATTGTATACCGC 353
           ||| | | | ||  | |||  |  ||  || | |  ||||||   ||| |    | | |
Db     645 GACAGCATCTGCATGAGATAGGATCATATTGTGAGCTGCAATGAAGGGTTCCCTTTCTGA 586

Qy     354 TTTCCCCGCAGGACAGTTGCATATTGTATTGGTCTATCGCTTCTCCCTGACTGGCTGAGC 413
           | ||| |||  ||| || |  | | ||   |   ||||  |      |   | | | |
Db     585 ATCCCCAGCACTACAATTCCCAAATATACCAGAGCATCGAGTTGGAGGGCAAGTCCCATC 526

Qy     414 TCTCT 418
           ||| |
Db     525 TCTAT 521


RESULT  10
US-08-670-186-1/c
; Sequence 1, Application US/08670186
; Patent No. 5859343
;  GENERAL INFORMATION:
;    APPLICANT:  SUN, SAMUEL S.M.
;    APPLICANT:  XIONG, LIWEN
;    APPLICANT:  HU, ZHONG
;    APPLICANT:  CHEN, HANG
;    TITLE OF INVENTION:  RECOMBINANT SWEET PROTEIN MABINLIN
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  2000 PENNSYLVANIA AVE NW, STE. 5500
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20006-1888
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US/08/670,186
;      FILING DATE:  21-JUN-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MURASHIGE, KATE H.
;      REGISTRATION NUMBER:  29,959
;      REFERENCE/DOCKET NUMBER:  23461-20007.00
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 887-1500
;      TELEFAX:  (202) 822-0168
;      TELEX:  90-4030 MRSNFOERSWSH
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  629 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  polyA_site
;      LOCATION:  629
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  28..492
;    FEATURE:
;      NAME/KEY:  mutation
;      LOCATION:  replace(229, "t")
;      OTHER INFORMATION:  /gene= "mutant A1"
;      OTHER INFORMATION:  /note= "codon changes from aspartic acid to tyrosine in mutan
;      OTHER INFORMATION:  A1"
;    FEATURE:
;      NAME/KEY:  mutation
;      LOCATION:  replace(230, "c")
;      OTHER INFORMATION:  /gene= "mutant A2"
;      OTHER INFORMATION:  /note= "codon changes from aspartic acid to alanine for mutan
;      OTHER INFORMATION:  A2"
;    FEATURE:
;      NAME/KEY:  mutation
;      LOCATION:  replace(272, "c")
;      OTHER INFORMATION:  /gene= "mutant L1"
;      OTHER INFORMATION:  /note= "codon changes from asparagine to threonine for mutant
;      OTHER INFORMATION:  L1"
;    FEATURE:
;      NAME/KEY:  mutation
;      LOCATION:  replace(272, "t")
;      OTHER INFORMATION:  /gene= "mutant L2"
;      OTHER INFORMATION:  /note= "codon changes from asparagine to isoleucine for mutan
;      OTHER INFORMATION:  L2"
US-08-670-186-1


  Query Match             4.7%;  Score 27;  DB 3;  Length 629;
  Best Local Similarity   47.9%;  Pred. No. 4.5;
  Matches   78;  Conservative    0;  Mismatches   85;  Indels    0;  Gaps    0;

Qy     322 TTTTCACATGCAATACCGGGATTGTATACCGCTTTCCCCGCAGGACAGTTGCATATTGTA 381
           ||  ||||    |    | |||| || |      ||| | | |           || ||
Db     560 TTAACACACTACAACATGCGATTCTACAAACACATCCTCTCCGTCAGCCACTGGATGGTT 501

Qy     382 TTGGTCTATCGCTTCTCCCTGACTGGCTGAGCTCTCTTTTTGGCTGTTGGTATGCACCTC 441
           | || |||  ||    | |||| |||   |||||   | |||| | |      |    |
Db     500 TCGGCCTAGGGCCATGCTCTGAATGGGCAAGCTCCGATGTTGGGTATGTTGCAGATGTTG 441

Qy     442 GCCAAAGGTGTAGCTCCCTCTGAAATAGGTGGTAATTGTTTTG 484
           | ||||  | | ||  | ||  | | | |    || || | ||
Db     440 GGCAAATTTCTTGCGGCATCGAAGAGACGCCTCAACTGCTGTG 398


RESULT  11
US-08-420-235B-1
; Sequence 1, Application US/08420235B
; Patent No. 5801042
;  GENERAL INFORMATION:
;    APPLICANT:  Chang, Yuan
;    APPLICANT:  Moore, Patrick S.
;    TITLE OF INVENTION:  UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
;    TITLE OF INVENTION:  SEQUENCES AND USES THEREOF
;    NUMBER OF SEQUENCES:  47
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/420,235B
;      FILING DATE:
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  45185-B
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  20710 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  N
;    ANTI-SENSE:  N
US-08-420-235B-1


  Query Match             4.6%;  Score 26.8;  DB 3;  Length 20710;
  Best Local Similarity   50.8%;  Pred. No. 37;
  Matches   64;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy      83 TGTTCCGTGAATCCGGAAATCTGCACGTACCTGCTCCAGATTACTATGAGGATTATCCTT 142
           | ||  ||   ||   ||   ||| |  |  | | ||| || |||| |||   |  |
Db  20257 TCTTTTGTCGTTCGTTAATCATGCCCTGAAGTACACCAAGTTTCTATACGGAGATTTTTC 20316

Qy     143 AAGTACAAGGGCCGCCGCCATCGTTCCGGTTCTTCCCACTCCGCCCAGACAATGAATCAT 202
           | ||| |||| || ||   |   |||    |  | ||| | |  | |||| |  |  |
Db  20317 CGGGACATGGGCGGCGGCTTGCCGCCCGCCATTCGCTACTTCTCGGATACAAAGGGTAGT 20376

Qy     203 CGGTAA 208
           || |
Db  20377 GAGTCA 20382


RESULT  12
US-08-420-235B-34
; Sequence 34, Application US/08420235B
; Patent No. 5801042
;  GENERAL INFORMATION:
;    APPLICANT:  Chang, Yuan
;    APPLICANT:  Moore, Patrick S.
;    TITLE OF INVENTION:  UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
;    TITLE OF INVENTION:  SEQUENCES AND USES THEREOF
;    NUMBER OF SEQUENCES:  47
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
```

```
;       ZIP:  10036
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/420,235B
;       FILING DATE:
;       CLASSIFICATION:  424
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  White, John P.
;       REGISTRATION NUMBER:  28,678
;       REFERENCE/DOCKET NUMBER:  45185-B
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 278-0400
;       TELEFAX:  (212) 391-0525
;   INFORMATION FOR SEQ ID NO:  34:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  984 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:  N
;     ANTI-SENSE:  N
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..984
;       OTHER INFORMATION:
US-08-420-235B-34


  Query Match             4.6%;  Score 26.8;  DB 3;  Length 984;
  Best Local Similarity   50.8%;  Pred. No. 6.8;
  Matches   64;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy      83 TGTTCCGTGAATCCGGAAATCTGCACGTACCTGCTCCAGATTACTATGAGGATTATCCTT 142
           | ||  ||   ||   ||   ||| |  |  | | ||| || ||||  |||   |  |
Db     597 TCTTTTGTCGTTCGTTAATCATGCCCTGAAGTACACCAAGTTTCTATACGGAGATTTTTC 656

Qy     143 AAGTACAAGGGCCGCCGCCATCGTTCCGGTTCTTCCCACTCCGCCCAGACAATGAATCAT 202
           | ||| |||| || ||   |   |||    |  | ||| | |  | |||| |  |  |
Db     657 CGGGACATGGGCGGCGGCTTGCCGCCCGCCATTCGCTACTTCTCGGATACAAAGGGTAGT 716

Qy     203 CGGTAA 208
           || |
Db     717 GAGTCA 722


RESULT  13
US-08-770-379-18
; Sequence 18, Application US/08770379
; Patent No. 5849564
;  GENERAL INFORMATION:
;    APPLICANT:  Chang, Yuan
;    APPLICANT:  Bohenzky, Roy A.
;    APPLICANT:  Russo, James J.
;    APPLICANT:  Edelman, Isidore S.
;    APPLICANT:  Moore, Patrick S.
;    TITLE OF INVENTION:  POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
;    TITLE OF INVENTION:  HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/770,379
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  52342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  35100 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-770-379-18


  Query Match             4.6%;  Score 26.8;  DB 3;  Length 35100;
  Best Local Similarity   50.8%;  Pred. No. 50;
  Matches   64;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy      83 TGTTCCGTGAATCCGGAAATCTGCACGTACCTGCTCCAGATTACTATGAGGATTATCCTT 142
           | ||  ||   ||   ||   ||| |  |  | | ||| || ||||  |||   |  |
Db  20171 TCTTTTGTCGTTCGTTAATCATGCCCTGAAGTACACCAAGTTTCTATACGGAGATTTTTC 20230

Qy     143 AAGTACAAGGGCCGCCGCCATCGTTCCGGTTCTTCCCACTCCGCCCAGACAATGAATCAT 202
           | ||| |||| || ||   |   |||    |  | ||| | |  | |||| |  |  |
Db  20231 CGGGACATGGGCGGCGGCTTGCCGCCCGCCATTCGCTACTTCTCGGATACAAAGGGTAGT 20290

Qy     203 CGGTAA 208
           || |
Db  20291 GAGTCA 20296


RESULT  14
PCT-US95-10194-1
; Sequence 1, Application PC/TUS9510194
;  GENERAL INFORMATION:
;    APPLICANT:  The Trustees of Columbia University in the City of New York
;    APPLICANT:  City
;    TITLE OF INVENTION:  UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
;    TITLE OF INVENTION:  SEQUENCES AND USES THEREOF
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/10194
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  45185-C-PCT/JPW/MSC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;  INFORMATION FOR SEQ ID NO:  1:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  20710 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  N
;    ANTI-SENSE:  N
PCT-US95-10194-1


  Query Match             4.6%;  Score 26.8;  DB 5;  Length 20710;
  Best Local Similarity   50.8%;  Pred. No. 37;
  Matches   64;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy      83 TGTTCCGTGAATCCGGAAATCTGCACGTACCTGCTCCAGATTACTATGAGGATTATCCTT 142
           | ||  ||   ||   ||   ||| |  |  | | ||| || ||||  |||   |  |
Db   20257 TCTTTTGTCGTTCGTTAATCATGCCCTGAAGTACACCAAGTTTCTATACGGAGATTTTTC 20316

Qy     143 AAGTACAAGGGCCGCCGCCATCGTTCCGGTTCTTCCCACTCCGCCCAGACAATGAATCAT 202
             | ||| |||| || ||   |   |||    |  | ||| | |  | |||| |  |  |
Db   20317 CGGGACATGGGCGGCGGCTTGCCGCCCGCCATTCGCTACTTCTCGGATACAAAGGGTAGT 20376

Qy     203 CGGTAA 208
             || |
Db   20377 GAGTCA 20382


RESULT  15
PCT-US95-10194-34
; Sequence 34, Application PC/TUS9510194
;  GENERAL INFORMATION:
;    APPLICANT:  The Trustees of Columbia University in the City of New York
;    APPLICANT:  City
;    TITLE OF INVENTION:  UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
;    TITLE OF INVENTION:  SEQUENCES AND USES THEREOF
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/10194
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  45185-C-PCT/JPW/MSC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;  INFORMATION FOR SEQ ID NO:  34:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  984 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  N
;    ANTI-SENSE:  N
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..984
;      OTHER INFORMATION:
```

```
PCT-US95-10194-34


  Query Match             4.6%;  Score 26.8;  DB 5;  Length 984;
  Best Local Similarity   50.8%;  Pred. No. 6.8;
  Matches   64;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy      83 TGTTCCGTGAATCCGGAAATCTGCACGTACCTGCTCCAGATTACTATGAGGATTATCCTT 142
           | ||  ||   ||   ||   ||| |  |  | | ||| || ||||  |||   |  |
Db     597 TCTTTTGTCGTTCGTTAATCATGCCCTGAAGTACACCAAGTTTCTATACGGAGATTTTTC 656

Qy     143 AAGTACAAGGGCCGCCGCCATCGTTCCGGTTCTTCCCACTCCGCCCAGACAATGAATCAT 202
             | ||| |||| || ||   |   |||    |  | ||| | |  | |||| |  |  |
Db     657 CGGGACATGGGCGGCGGCTTGCCGCCCGCCATTCGCTACTTCTCGGATACAAAGGGTAGT 716

Qy     203 CGGTAA 208
             || |
Db     717 GAGTCA 722


Search completed:  October  5, 1999, 18:01:16
Job time:  11516 sec
```

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 1999, 17:59:10 ; Search time 223.92 Seconds
(without alignments)
646.933 Million cell updates/sec

Title: US-09-068-804-10
Perfect score: 579
Sequence: 1 ATCAGCTTGCCGTCGTCATA.....CCAGCGCATCCCGGCGATTT 579

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	579	100.0	579	1	T67033	Salmonella tyrosin
2	82.4	14.2	2405	1	T37466	Salmonella enteric
3	31	5.4	434	1	T19501	Human gene signatu
4	31	5.4	323	1	V89846	EST clone CW1272.
c 5	30	5.2	10214	1	X20548	Polynucleotide seq
c 6	29.8	5.1	4992	1	V43720	Cancer associated
7	29.6	5.1	2071	1	T77131	Human RES cDNA. Ph
8	29.6	5.1	1098	1	V59891	DNA encoding Staph
9	29.6	5.1	5738	1	V74395	Staphylococcus aur
c 10	29.4	5.1	15464	1	Q11415	Ryanodine receptor
11	29.4	5.1	795	1	T67720	H. pylori inner me
12	29.4	5.1	1698	1	T68198	H. pylori inner me
c 13	29.2	5.0	80073	1	T58840_5	Continuation (6 of
c 14	29	5.0	1651	1	T97300	Streptococcus pneu
c 15	29	5.0	4694	1	V37372	Streptococcus pneu
16	29	5.0	16535	1	V52207	Streptococcus pneu
c 17	29	5.0	2573	1	V59153	Human ReproSA-1 co
18	29	5.0	4203	1	V64063	Zea mays promoter
c 19	29	5.0	1857	1	X07322	Streptococcus pneu
c 20	28.6	4.9	2319	1	Q73728	Chicken liver alph
c 21	28.6	4.9	5000	1	T29661	Nisin nisIPRK gene
c 22	28.6	4.9	7454	1	T29660	Nisin nisABTCIP ge
c 23	28.6	4.9	2334	1	T66121	Chicken alpha-N-ac
c 24	28.6	4.9	2319	1	T60489	Chicken liver alph
c 25	28.6	4.9	110000	1	V21209_12	Continuation (13 o
26	28.4	4.9	1692	1	T60974	NLERK2 cDNA 3' reg
c 27	28.4	4.9	22671	1	V31200	E. coli J96 pathog
c 28	28.4	4.9	7893	1	X13165	Enterococcus faeca
29	28.2	4.9	6416	1	Q24046	DNA sequence of pl
30	28.2	4.9	1113	1	Q62305	Clostridium perfri
31	28	4.8	1591	1	T93780	Oryza sativa patho
c 32	27.8	4.8	15672	1	Q10613	Rianodin receptor
c 33	27.8	4.8	15377	1	Q25975	MH mutant porcine
34	27.6	4.8	973	1	X24355	Wheat ATP-phosphor
c 35	27.4	4.7	24593	1	N50226	Sequence of opine
c 36	27.4	4.7	24596	1	N50182	Complete nucleotid
37	27.4	4.7	11800	1	Q67676	Acetyl CoA carboxy
c 38	27.4	4.7	1545	1	Q79100	Tobacco PMT PCR pr
39	27.4	4.7	12461	1	X32367	Human endothelin-1
40	27.2	4.7	110000	1	T58840_0	Mycoplasma genital
41	27.2	4.7	4164	1	V12216	Mouse retinoid met
c 42	27	4.7	2703	1	N90541	DNA encoding N-alp
c 43	27	4.7	2724	1	Q12226	N-alpha-acetyltran
44	27	4.7	6412	1	T28796	Human calcium sens
45	27	4.7	14141	1	X13331	Enterococcus faeca

ALIGNMENTS

RESULT 1
T67033
ID T67033 standard; DNA; 579 BP.
AC T67033;
DT 04-AUG-1997 (first entry)
DE Salmonella tyrosine phosphatase stpA gene.
KW StpA; tyrosine phosphatase; phosphorylation; signal transduction;
KW Salmonella secreted protein; Ssp; bacterial-mediated endocytosis;
KW ss.
OS Salmonella typhimurium.
FH Key Location/Qualifiers
FT cds 1. .579
FT /*tag= a
FT /product= StpA
FT /note= "the coding sequence does not translate
FT to the StpA amino acid sequence of Fig 25"
PN WO9718225-A1.
PD 22-MAY-1997.
PF 14-NOV-1996; U18504.
PR 14-NOV-1995; US-006733.
PA (GEHO) GEN HOSPITAL CORP.
PI Miller SI;
DR WPI; 97-289217/26.
PT New isolated Salmonella secreted proteins and related genes - used
PT to develop products for the detection, treatment or prevention of
PT Salmonella infections
PS Claim 15; Fig 24; 95pp; English.
CC Salmonella typhimurium stpA DNA (T67033) codes for Salmonella
CC tyrosine phosphatase (StpA) (W15273), an enzyme that catalyses the
CC release of phosphate groups from tyrosine residues in proteins.
CC StpA can be used to deactivate or activate eukaryotic or
CC prokaryotic signal transduction proteins that are regulated by the
CC phosphorylation and dephosphorylation of tyrosine residues, thereby
CC altering intracellular signal transduction. Thus, Stp can be used
CC as a research tool to study and evaluate phosphorylation-regulated
CC signal transduction pathways.
SQ Sequence 579 BP; 120 A; 158 C; 124 G; 177 T;

Query Match 100.0%; Score 579; DB 1; Length 579;
Best Local Similarity 100.0%; Pred. No. 4e-188;
Matches 579; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 ATCAGCTTGCCGTCGTCATAAGCAACTGGGCTTGCATTGCTTTTAGTTGTACAAACTGTG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATCAGCTTGCCGTCGTCATAAGCAACTGGGCTTGCATTGCTTTTAGTTGTACAAACTGTG 60

Qy     61 CAGGCGTCTTCCAGCATTCTATTGTTCCGTGAATCCGGAAATCTGCACGTACCTGCTCCA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CAGGCGTCTTCCAGCATTCTATTGTTCCGTGAATCCGGAAATCTGCACGTACCTGCTCCA 120

Qy    121 GATTACTATGAGGATTATCCTTAAGTACAAGGGCCGCCGCCATCGTTCCGGTTCTTCCCA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GATTACTATGAGGATTATCCTTAAGTACAAGGGCCGCCGCCATCGTTCCGGTTCTTCCCA 180

Qy    181 CTCCGCCCAGACAATGAATCATCGGTAAATGCTTATCTGATGAACTACGCCCCGGCGCGC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CTCCGCCCAGACAATGAATCATCGGTAAATGCTTATCTGATGAACTACGCCCCGGCGCGC 240

Qy    241 CATTTTGGTTACTATTTTTCACCCTATCCGCCAGGTATTCTAACTGATCCGTAGACGGTA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CATTTTGGTTACTATTTTTCACCCTATCCGCCAGGTATTCTAACTGATCCGTAGACGGTA 300

Qy    301 ACGGCTGGTGATCTGGCCAATTTTTCACATGCAATACCGGGATTGTATACCGCTTTCCCC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 ACGGCTGGTGATCTGGCCAATTTTTCACATGCAATACCGGGATTGTATACCGCTTTCCCC 360
```

```
Qy    361 GCAGGACAGTTGCATATTGTATTGGTCTATCGCTTCTCCCTGACTGGCTGAGCTCTCTTT 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GCAGGACAGTTGCATATTGTATTGGTCTATCGCTTCTCCCTGACTGGCTGAGCTCTCTTT 420

Qy    421 TTGGCTGTTGGTATGCACCTCGCCAAAGGTGTAGCTCCCTCTGAAATAGGTGGTAATTGT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TTGGCTGTTGGTATGCACCTCGCCAAAGGTGTAGCTCCCTCTGAAATAGGTGGTAATTGT 480

Qy    481 TTTGCCTGCATCTGATCTTCCGACGTTAACACCACCAGGCACGAGCATTCTTTTTCAAGA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TTTGCCTGCATCTGATCTTCCGACGTTAACACCACCAGGCACGAGCATTCTTTTTCAAGA 540

Qy    541 AGCATTTTCATATGCGCTTCCAGCGCATCCCGGCGATTT 579
          |||||||||||||||||||||||||||||||||||||||
Db    541 AGCATTTTCATATGCGCTTCCAGCGCATCCCGGCGATTT 579
```

```
RESULT   2
T37466
ID   T37466 standard; DNA; 2405 BP.
AC   T37466;
DT   25-FEB-1997  (first entry)
DE   Salmonella enterica IagA and IagB operon sequences.
KW   IagA; IagB; Salmonella enterica; primer; probe; HeLa; ss.
OS   Salmonella enterica.
FH   Key             Location/Qualifiers
FT   cds             97. .1758
FT                   /*tag=  a
FT                   /product= IagA protein
FT   misc_signal     1769. .1772
FT                   /*tag=  b
FT                   /note= "putative ribosome binding site"
FT   cds             1776. .2258
FT                   /*tag=  c
FT                   /product= IagB protein
PN   EP-721989-A1.
PD   17-JUL-1996.
PF   15-JAN-1996; 400098.
PR   16-JAN-1995; FR-000410.
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PA   (INSP ) INST PASTEUR.
PI   Le GUERN FELLOUS M,  Popoff MY;
DR   WPI; 96-322837/33.
DR   P-PSDB; W03528, W09225.
PT   New nucleic acid of S.enterica ssp. enterica involved in cell
PT   invasion  - and derived oligo-nucleotide(s) useful as primers and
PT   probes for detecting Salmonella in food etc.
PS   Claim 1; Fig 1; 33pp; French.
CC   Oligonucleotides T37451-65 are derived from the newly isolated sequence
CC   of the IagA and IagB sequences from Salmonella enterica ssp. enterica
CC   serovar typhi.  The IagA and IagB sequences (T37466) encode proteins
CC   which are involved in the invasion of cultured HeLa cells by S. enterica.
SQ   Sequence   2405 BP;    735 A;    488 C;    519 G;    663 T;
```

```
  Query Match             14.2%;  Score 82.4;  DB 1;  Length 2405;
  Best Local Similarity   97.9%;  Pred. No. 3.6e-18;
  Matches   94;  Conservative    0;  Mismatches    1;  Indels    1;  Gaps    1;

Qy      1 ATCAGCTTGCCGTCGTCATAAGCAACTGGGCTTGCATTGCTTTTAGTTGTACAAACTGTG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2311 ATCAGCTTGCCGTCGTCATAAGCAACTGGGCTTGCATTGCTTTTAGTTGTACAAACTGTG 2370

Qy     61 CAGGCGTCTTCCAGCATTCTATTGTTCCGTGAATCC 96
          |||||||||||||||||||||||||||||||| |
Db   2371 -AGGCGTCTTCCAGCATTCTATTGTTCCGTGAATTC 2405
```

```
RESULT   3
T19501
ID   T19501 standard; cDNA to mRNA; 434 BP.
AC   T19501;
DT   28-JUN-1996  (first entry)
DE   Human gene signature HUMGS00558.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 400; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   434 BP;    125 A;    72 C;    79 G;    151 T;
```

```
  Query Match             5.4%;  Score 31;  DB 1;  Length 434;
  Best Local Similarity   51.0%;  Pred. No. 0.56;
  Matches   73;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy    191 ACAATGAATCATCGGTAAATGCTTATCTGATGAACTACGCCCCGGCGCGCCATTTTGGTT 250
          || | |||||||  | | |  | | || ||||  | |||  |  |||||
Db     83 ACTAAGAATCATTCATGAGTAAATGTGTGTTGAATATCTACCCTTGACCTTTTTTTGAGA 142

Qy    251 ACTATTTTTCACCCTATCCGCCAGGTATTCTAACTGATCCGTAGACGGTAACGGCTGGTG 310
          | ||  | || |  |||     || |  | || ||  | |||  |  | ||  |
Db    143 AATAGAGTAAACACAGTCCCTGTAGTCTGACAGCTAATGGGGAGAGAGGGTAGACTTTTC 202

Qy    311 ATCTGGCCAATTTTTCACATGCA 333
          |||    || |||  |||||||
Db    203 ATCGAATTAAATTTCTACATGCA 225
```

```
RESULT   4
V89846
ID   V89846 standard; cDNA; 323 BP.
AC   V89846;
DT   15-FEB-1999  (first entry)
DE   EST clone CW1272.
KW   Human; secreted protein; expressed sequence tag; EST; haematopoiesis;
KW   tissue growth; activin; inhibin; chemotaxis; chemokinesis; haemostatic;
KW   receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;
KW   gene therapy; ss.
OS   Homo sapiens.
PN   WO9845436-A2.
PD   15-OCT-1998.
PF   10-APR-1998; U06955.
PR   10-APR-1997; US-838821.
PA   (GEMY ) GENETICS INST INC.
PI   Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
PI   Racie LA, Spaulding V, Treacy M;
DR   WPI; 99-070077/06.
```

PT New polynucleotides encoding human secreted proteins - derived from
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT ovary, pituitary, retina and colon cDNA libraries.
PS Claim 1; Page 349; 618pp; English.
CC The present sequence represents a human expressed sequence tag (EST).
CC The polynucleotide, which is a secreted EST, and the encoded protein
CC are predicted to have useful biological activities which would make
CC them suitable for treating, preventing or ameliorating medical
CC conditions in humans and animals, although no supporting data is
CC given. Suggested activities include nutritional activity, immune
CC stimulating or suppressing activity, haematopoiesis regulating
CC activity, tissue growth activity, activin/inhibin activity,
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic
CC activity, receptor/ligand activity, anti-inflammatory activity,
CC cadherin/tumour invasion suppressor activity, tumour inhibition
CC activity. The polynucleotide may also be useful for gene therapy.
SQ Sequence 323 BP; 60 A; 100 C; 95 G; 68 T;

Query Match 5.4%; Score 31; DB 1; Length 323;
Best Local Similarity 51.0%; Pred. No. 0.48;
Matches 73; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

```
Qy      50 TACAAACTGTGCAGGCGTCTTCCAGCATTCTATTGTTCCGTGAATCCGGAAATCTGCACG 109
           | | | ||||| |||| |  | | |||| ||  |||| |  |  |  | ||| |  | ||
Db      53 TGCCACCTGTGAAGGCCTTCTGCTGCATCCTCATGTTGGGGAAGATCAGAAGTGAGAACA 112

Qy     110 TACCTGCTCCAGATTACTATGAGGATTATCCTTAAGTACAAGGGCCGCCGCCATCGTTCC 169
           |      ||||     |   || |  |||  |  || || ||| |   ||| |||  |
Db     113 CATGCTGGCCAGTCCTTTCACAGCAGCATCAATCCGTGCATGGGGCAGGGCCCTCGGCCT 172

Qy     170 GGTTCTTCCCACTCCGCCCAGAC 192
           |      || |   ||||  ||
Db     173 GAGCGCCTCCCCCAGGCCCTAAC 195
```

RESULT 5
X20548/c
ID X20548 standard; DNA; 10214 BP.
AC X20548;
DT 05-MAY-1999 (first entry)
DE Polynucleotide sequence from the genome of Treponema pallidum.
KW Treponema pallidum infection; syphilis; Borrelia infection; animal;
KW enzyme production; ds.
OS Treponema pallidum.
PN WO9859034-A2.
PD 30-DEC-1998.
PF 23-JUN-1998; U13041.
PR 24-JUN-1997; US-050667.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Fraser CM;
DR WPI; 99-081273/07.
PT New isolated Treponema pallidum nucleic acids - used to develop
PT products for the detection, diagnosis, characterisation, prevention
PT and therapy of T. pallidum infections, particularly syphilis
PS Claim 1; Page 471-477; 1150pp; English.
CC X20500-21243 represent polynucleotide sequences from the genome of
CC Treponema pallidum. The sequences can be used for detection,
CC diagnosis, characterisation, prevention and therapy for T. pallidum
CC infections, particularly syphilis. They can also be used for detecting
CC diseases related to Borrelia infections in animals, and for the
CC production of biosynthetic products such as enzymes.
SQ Sequence 10214 BP; 2576 A; 3002 C; 2574 G; 2049 T;

Query Match 5.2%; Score 30; DB 1; Length 10214;
Best Local Similarity 55.9%; Pred. No. 5.8;
Matches 57; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

```
Qy     466 ATAGGTGGTAATTGTTTTGCCTGCATCTGATCTTCCGACGTTAACACCACCAGGCACGAG 525
           || | ||| ||||| ||| | ||| ||    |    || ||   ||| | | | ||
Db    2221 ATCGATGGCAATTGCTTTTTCAGCAGCTTCGGTCTGAGCGATACGGTCACGCGCCGCCAG 2162

Qy     526 CATTCTTTTTCAAGAAGCATTTTCATATGCGCTTCCAGCGCA 567
           || | ||||||  ||||| |   || |    ||   || ||
Db    2161 CACTTTTTTTCTCGAAGCGTCCGCAGACCGTATTAGCGCCCA 2120
```

RESULT 6
V43720/c
ID V43720 standard; cDNA to mRNA; 4992 BP.
AC V43720;
DT 16-NOV-1998 (first entry)
DE Cancer associated gene fragment 5.
KW ds; cancer; PCR; Northern blotting; ribonuclease protection assay;
KW diagnosis; metastatic cancer.
OS Synthetic.
PN WO9837187-A1.
PD 27-AUG-1998.
PF 18-FEB-1998; J00667.
PR 21-FEB-1997; JP-052508.
PA (TAKI) TAKARA SHUZO CO LTD.
PI Asada K, Hino F, Kato I, Mukai H, Yoshikawa Y;
DR WPI; 98-467552/40.
PT Detection of cancer cells in tissue samples - by changes in mRNA
PT expression compared to normal tissue of specific cancer-associated
PT gene sequences
PS Claim 1; Page 46-49; 92pp; Japanese.
CC The cancer associated gene fragments V43716-V43731 can be used to detect
CC cancer cells in tissue samples or biological fluids. They are detected
CC by monitoring the change in mRNA expression as compared to normal tissue
CC of one or more cancer-associated genes whose cDNA stringently hybridises
CC to the nucleic acid fragments. The change in expression may be an
CC increase or a decrease compared to normal tissue. The mRNA expression
CC may be determined by PCR, Northern blotting or ribonuclease protection
CC assay, or by determining the change in the amount of protein encoded by
CC the gene(s) as compared to normal tissue, for example by using a labelled
CC antibody recognising the protein. Detection of cancer cells for cancer
CC diagnosis, including detection of metastatic cancer cells in tissues
CC other than the primary tumour site.
SQ Sequence 4992 BP; 1705 A; 901 C; 1011 G; 1375 T;

Query Match 5.1%; Score 29.8; DB 1; Length 4992;
Best Local Similarity 49.7%; Pred. No. 4.8;
Matches 76; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

```
Qy     209 ATGCTTATCTGATGAACTACGCCCCGGCGCGCCATTTTGGTTACTATTTTTCACCCTATC 268
           ||| || | | ||  | |||   | |    | | ||   | |  || | ||||  | |
Db    1202 ATGATTTTTTAATTTATTACTAACAGTTTTCCAAGTTGCATGATAATCTCTCACAGTCTG 1143

Qy     269 CGCCAGGTATTCTAACTGATCCGTAGACGGTAACGGCTGGTGATCTGGCCAATTTTTCAC 328
           | | || | ||| |||| ||| || | |  | |       | ||      |    || |  |
Db    1142 CACAAGTTTTTCAAACTTATCTGTTGCTGTGACAACAAGTTGTCTCTGTGTTTTATAAGC 1083

Qy     329 ATGCAATACCGGGATTGTATACCGCTTTCCCCG 361
           || |||    || ||  | |   |||| |  ||
Db    1082 ATCCAAATAGGGAATAATTTTAGGCTTGCTACG 1050
```

RESULT 7
T77131
ID T77131 standard; cDNA; 2071 BP.
AC T77131;
DT 04-MAR-1998 (first entry)
DE Human RES cDNA.
KW Human PIG-F; phosphatidylinositol glucal class F; glycan; gene therapy;
KW tumour; RES; phenotype reversion; RRC6 cell line; v-src; RSV;
KW Rous sarcoma virus; viral kinase activity; ss.
OS Homo sapiens.
PN EP-785254-A2.
PD 23-JUL-1997.
PF 09-JAN-1997; 100216.
PR 18-JAN-1996; DE-001628.
PA (FARB) BAYER AG.
PI Ruebsamen-Waigmann H, Selders G, Strebhardt K;

```
DR   WPI; 97-365929/34.
PT   Phosphatidyl inositol glucal class F (PIG-F) and RES cDNA - useful
PT   in anti:sense orientation for gene therapy of tumours
PS   Claim 4; Pages 9-10; 10pp; German.
CC   This sequence represents the human RES cDNA. Res2.4 cDNA inserts have
CC   been shown to cause reversion of RRC6 cell line transformed phenotypes.
CC   RRC6 contains 5 copies of the v-src gene from RSV (Rous sarcoma virus)
CC   integrated into its cellular DNA and has constitutive viral kinase
CC   activity. A 306 bp stretch of RES cDNA 193 bp 3' untranslated region and
CC   113 bp open reading frame is identical to a part of a novel human PIG-F
CC   (phosphatidylinositol glucal class F) cDNA, which in antisense
CC   orientation is useful for gene therapy of tumours.
SQ   Sequence   2071 BP;    676 A;    372 C;    369 G;    654 T;

  Query Match              5.1%;  Score 29.6;  DB 1;  Length 2071;
  Best Local Similarity   59.5%;  Pred. No. 3.6;
  Matches   50;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy    457 CCCTCTGAAATAGGTGGTAATTGTTTTGCCTGCATCTGATCTTCCGACGTTAACACCACC 516
          | ||||||| ||| |||||||| ||||   |  ||     |||  |  ||  ||| ||
Db    859 CTCTCTGAACTAGTTGGTAATTTTTTTTTTTTAATTCCCACTTTGGCTGTGTACATCAAA 918

Qy    517 AGGCACGAGCATTCTTTTTCAAGA 540
            |  | ||| | | | |||   ||
Db    919 TGAAATGAGAAGTGTGTTTGCTGA 942


RESULT   8
V59891
ID   V59891 standard; DNA; 1098 BP.
AC   V59891;
DT   24-NOV-1998  (first entry)
DT   27-NOV-1998  (revised)
DE   DNA encoding Staphylococcus aureus proteins.
KW   Treatment; prevention; bacterial infection; Helicobacter pylori;
KW   vaccine; ss.
OS   Staphylococcus aureus.
PN   WO9823738-A2.
PD   04-JUN-1998.
PF   24-NOV-1997; U22092.
PR   25-NOV-1996; US-031469.
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PI   Warren RL;
DR   WPI; 98-322718/28.
PT   New nucleic acid from Staphylococcus aureus NCIMB 40771 - useful
PT   for, e.g. diagnosis, prevention and treatment of bacterial
PT   infection(s)
PS   Claim 4; Pages 61-62; 114pp; English.
CC   The present sequence is obtained from Staphylococcus aureus WCHU (NCIMB
CC   40771). The specification describes S. aureus proteins which are used to
CC   generate antibodies and to screen for antimicrobials. The products are
CC   used to treat or prevent bacterial infections, particularly
CC   where caused by S. aureus but also against Helicobacter pylori.
CC   Particular applications are to treat subjects before surgery or
CC   insertion of an in-dwelling device (alternatively the device itself is
CC   impregnated before placement). The nucleic acid sequence is used as
CC   sources of antisense sequences (for therapeutic use) or regulatory
CC   elements for controlling expression of bacterial genes, and for
CC   antibacterial screening. The protein can be also used as a vaccine.
CC   note: record reissued to correct DR PSDB line.
SQ   Sequence   1098 BP;    371 A;    150 C;    223 G;    353 T;

  Query Match              5.1%;  Score 29.6;  DB 1;  Length 1098;
  Best Local Similarity   64.7%;  Pred. No. 2.7;
  Matches   44;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy     11 CGTCGTCATAAGCAACTGGGCTTGCATTGCTTTTAGTTGTACAAACTGTGCAGGCGTCTT 70
          ||| |  ||| || | || |  || |||| ||||| ||||  || |||||  | | |||
Db    175 CGTGGGAATATGCCAGTGTGAGTGAATTGATTTTAATTGTTAAAGCTGTGACGACATCTA 234

Qy     71 CCAGCATT 78
          |  |||
Db    235 TCGTTATT 242


RESULT   9
V74395
ID   V74395 standard; DNA; 5738 BP.
AC   V74395;
DT   16-MAR-1999  (first entry)
DE   Staphylococcus aureus contig SEQ ID #84.
KW   Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW   cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW   skin infection; surgical wound infection; scalded skin syndrome;
KW   toxic shock syndrome; ds.
OS   Staphylococcus aureus.
FH   Key             Location/Qualifiers
FT   misc_feature    1441. .1500
FT                   /*tag= a
FT                   /note= "these bases represent a line of missing text in
FT                           the sequence listing in the specification. They
FT                           are included to maintain the nucleotide numbering
FT                           given in the specification for this DNA sequence"
FT   misc_feature    3241. .3300
FT                   /*tag= b
FT                   /note= "these bases represent a line of missing text in
FT                           the sequence listing in the specification. They
FT                           are included to maintain the nucleotide numbering
FT                           given in the specification for this DNA sequence"
FT   misc_feature    5041. .5100
FT                   /*tag= c
FT                   /note= "these bases represent a line of missing text in
FT                           the sequence listing in the specification. They
FT                           are included to maintain the nucleotide numbering
FT                           given in the specification for this DNA sequence"
PN   EP-786519-A2.
PD   30-JUL-1997.
PF   07-JAN-1997; 100117.
PR   05-JAN-1996; US-009861.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI   Rosen CA;
DR   WPI; 97-374922/35.
PT   Polynucleotide(s) and proteins derived from Staphylococcus aureus -
PT   stored on computer readable medium and used in the production of
PT   anti-S.aureus vaccines
PS   Claim 1; Page 541-544; 3271pp; English.
CC   This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC   of the invention. The DNA sequences are recorded on a computer readable
CC   medium, preferably selected from a floppy or hard disk, random access
CC   memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC   the S.aureus DNA sequences allows putative functions to be assigned so
CC   that protein-encoding or regulatory regions of commercial, therapeutic or
CC   industrial importance can be obtained. Specifically, sequences which are
CC   likely to encode antigens have been identified and these polypeptides can
CC   be used in a vaccine composition against S.aureus infection. The
CC   polypeptides can also be used in a kit for the immunodetection of
CC   S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC   including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC   skin and surgical wound infections, scalded skin syndrome, toxic shock
CC   syndrome, etc. Organisms transformed with the DNA sequences can be used
CC   for recombinant production of the polypeptides. The new DNA sequences
CC   (and their fragments) are useful as primers or probes for isolating
CC   homologues of any of the S.aureus DNA sequences contained on the
CC   computer readable medium.
SQ   Sequence   5738 BP;    2014 A;    776 C;    1116 G;    1648 T;

  Query Match              5.1%;  Score 29.6;  DB 1;  Length 5738;
  Best Local Similarity   64.7%;  Pred. No. 6;
  Matches   44;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy     11 CGTCGTCATAAGCAACTGGGCTTGCATTGCTTTTAGTTGTACAAACTGTGCAGGCGTCTT 70
          ||| |  ||| || | || |  || |||| ||||| ||||  || |||||  | | |||
```

```
Db   1882 CGTGGGAATATGCCAGTGTGAGTGAATTGATTTTAATTGTTAAAGCTGTGACGACATCTA 1941

Qy     71 CCAGCATT 78
          |   |||
Db   1942 TCGTTATT 1949


RESULT  10
Q11415/c
ID   Q11415 standard; cDNA; 15464 BP.
AC   Q11415;
DT   17-JUN-1991  (first entry)
DE   Ryanodine receptor gene.
KW   Malignant hyperthermia; hypermetabolic syndrome; inhalation;
KW   anaesthetics; probe; calcium release channel; sarcoplasmic;
KW   reticulum; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             105. .15319
FT                   /*tag= a
FT                   /product= ryanodine receptor
FT   polya_signal    15439. .15446
FT                   /*tag= b
FT   misc_rna        15447. .15457
FT                   /*tag= c
FT                   /label= TG-rich region
FT   polya_site      15464
FT                   /*tag= d
FT   misc_rna        328. .773
FT                   /*tag= e
FT                   /label= HRR-7B
FT                   /note= "0.4 kb; claim 6"
FT   misc_rna        400. .2620
FT                   /*tag= f
FT                   /label= HRR-7A
FT                   /note= "2.2 kb; claim 6"
FT   misc_rna        2615. .6359
FT                   /*tag= g
FT                   /label= HRR-5
FT                   /note= "3.8 kb; claim 6"
FT   misc_rna        6349. .8717
FT                   /*tag= h
FT                   /label= HRR-4
FT                   /note= "2.4 kb; claim 6"
FT   misc_rna        9773. .11075
FT                   /*tag= i
FT                   /label= HRR-3-1200
FT                   /note= "1.2 kb; claim 6"
FT   misc_rna        8738. .9778
FT                   /*tag= j
FT                   /label= HRR-3-1000
FT                   /note= "1.0 kb; claim 6"
FT   misc_rna        8739. .11842
FT                   /*tag= k
FT                   /label= HRR-3
FT                   /note= "3.1 kb; claim 6"
FT   misc_rna        11837. .13831
FT                   /*tag= l
FT                   /label= HRR-2
FT                   /note= "2.0 kb; claim 6"
FT   misc_rna        13826. .15467
FT                   /*tag= m
FT                   /label= HRR-1
FT                   /note= "1.6 kb; claim 6"
PN   WO9104328-A.
PD   04-APR-1991.
PF   21-SEP-1990; CA0312.
PR   25-SEP-1989; US-612726.
PA   (HSCR-) HSC R & D Partnership.
PA   (UTOR) Univ. Toronto Innova.
PA   (TORO-) Toronto Hospital.
PI   Worton RG, Maclennan DH, Britt BA;
DR   WPI; 91-117517/16.
DR   P-PSDB; R11510.
PT   Purified DNA specific for human ryanodine receptor - useful for
PT   diagnosis of malignant melanoma.
PS   Claim 1; Fig 2; 49pp; English.
CC   The sequence was obtd. from several overlapping clones isolated
CC   from a human skeletal muscle cDNA library in lambda gt10. The gene
CC   (located on chromosome 19) encodes the human ryanodine receptor
CC   (HRR), a calcium release channel which spans the gap between the
CC   transverse tubule and the sarcoplasmic reticulum (SR) in muscle.
CC   The sustained muscle contraction in malignant hyperthermia (MH)
CC   may be caused by the release of calcium into the muscle cell cyto-
CC   plasm from the SR.  This is due to a defect in the gene encoding
CC   HRR.  Sequences from the gene can be used for diagnosis of MH
CC   using FRPL analysis (see tags e-m).
SQ   Sequence   15464 BP;    3286 A;    4616 C;    4735 G;    2827 T;


  Query Match             5.1%;  Score 29.4;  DB 1;  Length 15464;
  Best Local Similarity   70.9%;  Pred. No. 11;
  Matches   39;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy    426 TGTTGGTATGCACCTCGCCAAAGGTGTAGCTCCCTCTGAAATAGGTGGTAATTGT 480
          ||||||  | ||||| |  |||||||| |||| || ||  || ||  ||||  ||
Db   4734 TGTTGGGTTCCACCTGGAAAAAGGTGTTGCTCTCTTTGCCATTGGCTGTAAAGGT 4680


RESULT  11
T67720
ID   T67720 standard; DNA; 795 BP.
AC   T67720;
DT   17-JUL-1997  (first entry)
DE   H. pylori inner membrane protein ORF 9776562.aa.
KW   Cytoplasmic; vaccine; prevention; treatment; infection; identification;
KW   binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW   duodenal ulcer disease; chronic gastritis; diagnosis; envelope; ds.
OS   Helicobacter pylori.
FH   Key             Location/Qualifiers
FT   cds             1. .795
FT                   /*tag= a
FT                   /transl_except= (pos: 757. .759, aa: Xaa)
FT                   /note= "Xaa = unknown; no stop codon given"
PN   WO9640893-A1.
PD   19-DEC-1996.
PF   06-JUN-1996; U09122.
PR   07-JUN-1995; US-487032.
PR   01-APR-1996; US-630405.
PA   (ASTR ) ASTRA AB.
PI   Berglindh OT,  Smith D,  Mellgaerd BL;
DR   WPI; 97-052306/05.
DR   P-PSDB; W20580.
PT   Helicobacter pylori nucleic acid sequences and related
PT   polypeptide(s)  - useful for vaccines to treat or prevent H. pylori
PT   infection, and to detect Helicobacter
PS   Claim 1; Page 320; 1481pp; English.
CC   This sequence encodes a H. pylori inner membrane protein.
CC   The protein may be used in a vaccine to prevent or treat H. pylori
CC   infection or to identify H. pylori polypeptide binding compounds,
CC   useful as potential H. pylori life cycle activators or inhibitors.
CC   The genomic sequence of H. pylori (ATCC 55679) was determined from
CC   overlapping contigs generated by mechanically shearing the bacterial
CC   DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC   and the predicted coding regions defined by computer evaluation. To
CC   identify likely H. pylori antigens for vaccine development, the amino
CC   acid sequences predicted from various ORF were analysed for significant
CC   homology to other known or exported membrane proteins. Having identified
CC   and determined the sequences of interest, particular regions can be
CC   isolated from H. pylori by PCR amplification for recombinant polypeptide
CC   production, e.g. in E. coli hosts.
SQ   Sequence   795 BP;    219 A;    142 C;    151 G;    282 T;


  Query Match             5.1%;  Score 29.4;  DB 1;  Length 795;
  Best Local Similarity   54.1%;  Pred. No. 2.7;
```

```
  Matches   60;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy    438 CCTCGCCAAAGGTGTAGCTCCCTCTGAAATAGGTGGTAATTGTTTTGCCTGCATCTGATC 497
          | | ||||   | ||| |||    ||||| |  |   |  |   ||   |||| |
Db     18 CTTTGCCAGTAGCGTAACTCTAGATGAAAAATTTTTAATGTTCCTTCTTTGCAACGCTCT 77

Qy    498 TTCCGACGTTAACACCACCAGGCACGAGCATTCTTTTTCAAGAAGCATTTT 548
          |||  ||| | |||  |  ||  |   |  |||||| || | | || ||||
Db     78 TTCTAACGCTTACAAAAATAGCGATTTGTTTTCTTTCTCTAAAGGCTTTTT 128


RESULT  12
T68198
ID   T68198 standard; DNA; 1698 BP.
AC   T68198;
DT   21-JUL-1997  (first entry)
DE   H. pylori inner membrane protein ORF 29zp10241orf14.
KW   Vaccine; prevention; treatment; infection; identification;
KW   binding compound; bacterium; life cycle; activator; bacteria;
KW   inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;
KW   inner membrane; ds.
OS   Helicobacter pylori.
FH   Key             Location/Qualifiers
FT   cds             1. .1698
FT                   /*tag=  a
FT                   /note= "no stop codon given"
PN   WO9640893-A1.
PD   19-DEC-1996.
PF   06-JUN-1996; U09122.
PR   07-JUN-1995; US-487032.
PR   01-APR-1996; US-630405.
PA   (ASTR ) ASTRA AB.
PI   Berglindh OT,  Smith D,  Mellgaerd BL;
DR   WPI; 97-052306/05.
DR   P-PSDB; W20945.
PT   Helicobacter pylori nucleic acid sequences and related
PT   polypeptide(s)  - useful for vaccines to treat or prevent H. pylori
PT   infection, and to detect Helicobacter
PS   Claim 1; Pages 958-959; 1481pp; English.
CC   The present sequence encodes a Helicobacter pylori inner membrane
CC   protein.
CC   The protein may be used in a vaccine to prevent or treat H. pylori
CC   infection or to identify H. pylori polypeptide binding compounds,
CC   useful as potential H. pylori life cycle activators or inhibitors.
CC   The genomic sequence of H. pylori (ATCC 55679) was determined from
CC   overlapping contigs generated by mechanically shearing the
CC   bacterial DNA. The sequences were analysed for ORF of at least 180
CC   nucleotides, and the predicted coding regions defined by computer
CC   evaluation. To identify likely H. pylori antigens for vaccine
CC   development, the amino acid sequences predicted from various ORF
CC   were analysed for significant homology to other known or exported
CC   membrane proteins. Having identified and determined the sequences
CC   of interest, particular regions can be isolated from H. pylori by
CC   PCR amplification for recombinant polypeptide production, e.g. in
CC   E. coli hosts.
SQ   Sequence   1698 BP;    515 A;    309 C;    315 G;    559 T;


  Query Match             5.1%;  Score 29.4;  DB 1;  Length 1698;
  Best Local Similarity   54.1%;  Pred. No. 3.8;
  Matches   60;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy    438 CCTCGCCAAAGGTGTAGCTCCCTCTGAAATAGGTGGTAATTGTTTTGCCTGCATCTGATC 497
          | | ||||   | ||| |||    ||||| |  |   |  |   ||   |||| |
Db     57 CTTTGCCAGTAGCGTAACTCTAGATGAAAAATTTTTAATGTTCCTTCTTTGCAACGCTCT 116

Qy    498 TTCCGACGTTAACACCACCAGGCACGAGCATTCTTTTTCAAGAAGCATTTT 548
          |||  ||| | |||  |  ||  |   |  |||||| || | | || ||||
Db    117 TTCTAACGCTTACAAAAATAGCGATTTGTTTTCTTTCTCTAAAGGCTTTTT 167


RESULT  13
T58840_5/c
```

```
Continuation (6 of 6) of T58840 from base 500001 (Mycoplasma genitalium genome. )
WP Sequence split into 6 fragments  LOCUS T58840 Accession T58840
WP      Fragment Name       Begin      End
WP      T58840_0                1     110000
WP      T58840_1           100001     210000
WP      T58840_2           200001     310000
WP      T58840_3           300001     410000
WP      T58840_4           400001     510000
WP      T58840_5           500001     580073


  Query Match             5.0%;  Score 29.2;  DB 1;  Length 80073;
  Best Local Similarity   65.2%;  Pred. No. 30;
  Matches   43;  Conservative    0;  Mismatches   23;  Indels    0;  Gaps    0;

Qy    274 GGTATTCTAACTGATCCGTAGACGGTAACGGCTGGTGATCTGGCCAATTTTTCACATGCA 333
          ||||   ||| | ||   |  || | || |  ||||||  ||||||  |  |  | |||||
Db  34575 GGTAAATTAAATAATGTTTGAACTGGAATGTGTGGTGAAATGGCCAGTGATCAATATGCA 34516

Qy    334 ATACCG 339
          ||||||
Db  34515 ATACCG 34510


RESULT  14
T97300/c
ID   T97300 standard; DNA; 1651 BP.
AC   T97300;
DT   28-APR-1998  (first entry)
DE   Streptococcus pneumoniae prolyl tRNA synthetase gene.
KW   Prolyl tRNA synthetase; proS polypeptide; vaccine;
KW   genetic immunisation; antibacterial; antibiotic; otitis media;
KW   conjunctivitis; pneumonia; bacteraemia; meningitis; sinusitis;
KW   pleural emphysema; endocarditis; gene therapy; ss.
OS   Streptococcus pneumoniae.
FH   Key             Location/Qualifiers
FT   CDS             1. .1650
FT                   /*tag=  a
FT                   /product= "proS"
FT                   /note= "no stop codon given"
PN   WO9738714-A1.
PD   23-OCT-1997.
PF   18-APR-1997; U06446.
PR   18-APR-1996; GB-007991.
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PA   (SMIK ) SMITHKLINE BEECHAM PLC.
PI   Lawlor EJ;
DR   WPI; 97-526207/48.
DR   P-PSDB; W31585.
PT   DNA encoding prolyl-tRNA synthetase from Streptococcus pneumoniae  -
PT   useful in the production of vaccines for protection against
PT   bacterial infection
PS   Claim 5; Page 34; 44pp; English.
CC   This isolated polynucleotide comprises a full-length coding sequence
CC   for prolyl tRNA synthetase (proS) of Streptococcus pneumoniae. The
CC   polynucleotide, or a claimed fragment of it (see T97301), can be used
CC   in a process for producing proS polypeptide in a host cell. ProS or its
CC   coding sequence (through gene therapy) are used to induce an
CC   immunological response in a mammal to generate antibodies and/or T cell
CC   immune response to protect against disease, especially otitis media,
CC   conjunctivitis, pneumonia, bacteraemia, meningitis, sinusitis, pleural
CC   emphysema and endocarditis. Antibodies against proS protect against
CC   invasion of bacteria, e.g. by blocking adherence of bacteria to damaged
CC   tissue, including wounds in skin or connective tissue caused by
CC   mechanical, chemical or thermal damage of by implantation of in-dwelling
CC   devices, or wounds in the mucous membranes. Antagonists are used to
CC   inhibit the proS protein.
SQ   Sequence   1651 BP;    479 A;    326 C;    388 G;    458 T;


  Query Match             5.0%;  Score 29;  DB 1;  Length 1651;
  Best Local Similarity   57.0%;  Pred. No. 5.2;
  Matches   53;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;
```

```
Qy    369 GTTGCATATTGTATTGGTCTATCGCTTCTCCCTGACTGGCTGAGCTCTCTTTTTGGCTGT 428
          | |||  ||| | | ||   || | ||| ||  |  |  | |   || | |||||||  |
Db   1562 GGTGCCAATTCTTTAGGGAAATTGATTCCCCAAGCGTAACGGTATTCACCTTTTGGCGTT 1503

Qy    429 TGGTATGCACCTCGCCAAAGGTGTAGCTCCCTC 461
          | ||   ||    | | |  |||| ||||| ||
Db   1502 TTGTTAACAAAGAGGCGAGCGTGTTGCTCCATC 1470


RESULT 15
V37372/c
ID   V37372 standard; DNA; 4694 BP.
AC   V37372;
DT   13-OCT-1998  (first entry)
DE   Streptococcus pneumoniae coding region.
KW   coding region; ORF; open reading frame; antibacterial;
KW   infection; prevention; meningitis; ss.
OS   Streptococcus pneumoniae.
FH   Key             Location/Qualifiers
FT   CDS             212. .940
FT                   /*tag=  a
FT                   /product= unknown
FT   CDS             1202. .1753
FT                   /*tag=  b
FT                   /product= putative cdsA homologue
FT   CDS             2750. .3037
FT                   /*tag=  c
FT                   /product= unknown
PN   WO9819689-A1.
PD   14-MAY-1998.
PF   27-OCT-1997; U19226.
PR   01-NOV-1996; US-029930.
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PA   (SMIK ) SMITHKLINE BEECHAM PLC.
PI   Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO,
PI   Reid RH, Zarfos PN;
DR   WPI; 98-286586/25.
DR   P-PSDB; W60977, W60978, W60979.
PT   New isolated nucleic acids from Streptococcus pneumoniae - useful,
PT   e.g. for identifying anti-bacterial(s) for treatment and prevention
PT   of meningitis
PS   Claim 1; Page 72-74; 130pp; English.
CC   The sequence is that of a coding region isolated from
CC   S. pneumoniae. Its encoded protein, or agonists of it,
CC   may be useful as an antibacterial for treatment or
CC   prevention of infection, specifically caused by S. pneumoniae
CC   (particularly meningitis) but possibly also Helicobacter
CC   pylori (ulcers and gastric cancer). It may be of particular
CC   use before insertion of an in-dwelling device or any other
CC   invasive procedure. The protein, or nucleic acid encoding
CC   it, can also be used in vaccines to induce a cellular
CC   and/or humoral immune response, or to screen for other
CC   antibacterials. The DNA may also contain flanking sequences
CC   that are potential sources of control elements for bacterial
CC   gene expression. Detecting a sequence encoding the protein
CC   can be used diagnostically, e.g. to detect a mutation for
CC   serotyping or classifying infectious agents.
SQ   Sequence   4694 BP;    1312 A;    881 C;    1103 G;    1395 T;


  Query Match             5.0%;  Score 29;  DB 1;  Length 4694;
  Best Local Similarity   57.0%;  Pred. No. 8.7;
  Matches   53;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy    369 GTTGCATATTGTATTGGTCTATCGCTTCTCCCTGACTGGCTGAGCTCTCTTTTTGGCTGT 428
          | |||  ||| | | ||   || | ||| ||  |  |  | |   || | |||||||  |
Db   4608 GGTGCCAATTCTTTAGGGAAATTGATTCCCCAAGCGTAACGGTATTCACCTTTTGGCGTT 4549

Qy    429 TGGTATGCACCTCGCCAAAGGTGTAGCTCCCTC 461
          | ||   ||    | | |  |||| ||||| ||
Db   4548 TTGTTAACAAAGAGGCGAGCGTGTTGCTCCATC 4516
```

Search completed: October 5, 1999, 17:59:29
Job time: 11411 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 1999, 18:02:15 ; Search time 2414.69 Seconds
(without alignments)
762.582 Million cell updates/sec

Title: US-09-068-804-10
Perfect score: 579
Sequence: 1 ATCAGCTTGCCGTCGTCATA.....CCAGCGCATCCCGGCGATTT 579

Scoring table: IDENTITY_NUC

Searched: 679419 seqs, 1590154680 residues

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_st:*
14: gb_sts:*
15: gb_sy:*
16: gb_un:*
17: gb_vi:*
18: em_fun:*
19: em_htg:*
20: em_hum1:*
21: em_hum2:*
22: em_in:*
23: em_om:*
24: em_or:*
25: em_ov:*
26: em_pat:*
27: em_ph:*
28: em_pl:*
29: em_ro:*
30: em_sts:*
31: em_sy:*
32: em_un:*
33: em_vi:*
34: gb_htg1:*
35: gb_htg2:*
36: gb_in1:*
37: gb_in2:*
38: em_ba1:*
39: em_ba2:*
40: em_hum3:*
41: em_hum4:*
42: gb_pr4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	526	90.8	1691	2	STU63293	U63293 Salmonella
c 2	519.6	89.7	3964	1	SISIPSTP	X92546 S.typhi sip
3	82.4	14.2	2406	1	STIAGAB	X80892 S.typhi gen
4	82.4	14.2	2405	5	A52315	A52315 Sequence 19
5	48.6	8.4	7218	5	I66494	I66494 Sequence 14
c 6	34.2	5.9	90859	8	ATAC005561	AC005561 Arabidops
c 7	34.2	5.9	162346	35	AC006415	AC006415 Drosophil
c 8	33.6	5.8	1318	10	HSU90917	U90917 Human clone
c 9	33.6	5.8	259037	35	AC005911	AC005911 Homo sapi
10	33.6	5.8	97078	42	AC005841	AC005841 Homo sapi
11	33.4	5.8	14200	36	CEW07E11	Z49868 Caenorhabdi
c 12	33	5.7	84981	35	AC005472	AC005472 Drosophil
c 13	33	5.7	3816	36	S72379	S72379 vg=vestigia
14	32.8	5.7	1369	4	DRU34590	U34590 Danio rerio
15	32.8	5.7	203300	42	AC000134	AC000134 Homo sapi
c 16	32.2	5.6	35980	1	MTY20H10	Z92772 Mycobacteri
17	31.8	5.5	63749	11	AC004694	AC004694 Homo sapi
c 18	31.8	5.5	150070	35	AC007720	AC007720 Homo sapi
c 19	31.8	5.5	85579	42	AC004744	AC004744 Homo sapi
20	31.6	5.5	244254	10	HSAC001228	AC001228 244Kb Con
c 21	31.6	5.5	37972	10	HSU51281	U51281 Human chrom
c 22	31.6	5.5	141475	11	AC002980	AC002980 Homo sapi
23	31.4	5.4	272125	35	AC006513	AC006513 Homo sapi
24	31.2	5.4	2426	4	GGACTI	V01507 Gallus gall
c 25	31.2	5.4	106988	10	AC001226	AC001226 Genomic s
26	31.2	5.4	9608	17	YMU42596	U42596 Yam mosaic
27	31	5.4	52232	7	AB028617	AB028617 Arabidops
28	31	5.4	157248	11	AC002312	AC002312 Human Chr
29	31	5.4	134929	11	AC004655	AC004655 Homo sapi
30	31	5.4	100259	11	AC005800	AC005800 Homo sapi
31	31	5.4	185286	12	AC006508	AC006508 Mus muscu
32	31	5.4	434	14	G26168	G26168 human STS E
33	31	5.4	434	14	G27130	G27130 human STS S
c 34	31	5.4	255362	35	AC006401	AC006401 Mus muscu
c 35	31	5.4	231894	35	AC007221	AC007221 Homo sapi
c 36	30.8	5.3	41293	34	CEH06C16	Z92791 Caenorhabdi
c 37	30.8	5.3	195887	34	CNS00007	AL049780 Homo sapi
38	30.8	5.3	203010	42	AC006530	AC006530 Homo sapi
c 39	30.6	5.3	1553	1	CLORG16SAB	L09168 Clostridium
c 40	30.6	5.3	134199	1	SYCSLRF	D64004 Synechocyst
c 41	30.6	5.3	2310	7	FNU16956	U16956 Filobasidie
c 42	30.6	5.3	111906	8	IG005I10	AF013293 Arabidops
43	30.6	5.3	105127	11	AC004548	AC004548 Homo sapi
44	30.6	5.3	172366	11	AC004949	AC004949 Homo sapi
c 45	30.6	5.3	110000	34	CEY57G11_0	Z92841 Caenorhabdi

ALIGNMENTS

RESULT 1
STU63293/c
LOCUS STU63293 1691 bp DNA BCT 05-SEP-1996
DEFINITION Salmonella typhimurium protein tyrosine phosphatase SptP (SptP) gene, complete cds.
ACCESSION U63293
NID g1519053
VERSION U63293.1 GI:1519053
KEYWORDS .
SOURCE Salmonella typhimurium.
ORGANISM Salmonella typhimurium
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Salmonella.
REFERENCE 1 (bases 1 to 1691)
AUTHORS Kaniga,K., Uralil,J., Bliska,J.B. and Galan,J.E.
TITLE A secreted protein tyrosine phosphatase with modular effector domains in the bacterial pathogen Salmonella typhimurium
JOURNAL Mol. Microbiol. 21 (3), 633-641 (1996)
MEDLINE 97020031
REFERENCE 2 (bases 1 to 1691)
AUTHORS Galan,J.E.
TITLE Direct Submission
JOURNAL Submitted (09-JUL-1996) Molecular Genetics and Microbiology, State University of New York at Stony Brook, Stony Brook, NY 11794-522,

```
          USA
FEATURES             Location/Qualifiers
     source          1. .1691
                     /organism="Salmonella typhimurium"
                     /strain="SL1344"
                     /db_xref="taxon:602"
     gene            11. .1642
                     /gene="SptP"
     CDS             11. .1642
                     /gene="SptP"
                     /note="involved in the virulence of Salmonella
                     typhimurium"
                     /codon_start=1
                     /transl_table=11
                     /product="protein tyrosine phosphatase SptP"
                     /protein_id="AAC44349.1"
                     /db_xref="PID:g1519054"
                     /db_xref="GI:1519054"
                     /translation="MLKYEERKLNNLTLSSFSKVGVSNDARLYIAKENTDKAYVAPEK
                     FSSKVLTWLGKMPLFKNTEVVQKHTENIRVQDQKILQTFLHALTEKYGETAVNDALLM
                     SRINMNKPLTQRLAVQITECVKAADEGFINLIKSKDNVGVRNAALVIKGGDTKVAEKN
                     NDVGAESKQPLLDIALKGLKRTLPQLEQMDGNSLRENFQEMASGNGPLRSLMTNLQNL
                     NKIPEAKQLNDYVTTLTNIQVGVARFSQWGTCGGEVERWVDKASTHELTQAVKKIHVI
                     AKELKNVTAELEKIEAGAPMPQTMSGPTLGLARFAVSSIPINQQTQVKLSDGMPVPVN
                     TLTFDGKPVALAGSYPKNTPDALEAHMKMLLEKECSCLVVLTSEDQMQAKQLPPYFRG
                     SYTFGEVHTNSQKVSSASQGEAIDQYNMQLSCGEKRYTIPVLHVKNWPDHQPLPSTDQ
                     LEYLADRVKNSNQNGAPGRSSSDKHLPMIHCLGGVGRTGTMAAALVLKDNPHSNLEQV
                     RADFRDSRNNRMLEDASQFVQLKAMQAQLLMTTAS"
BASE COUNT      562 a    326 c    412 g    391 t
ORIGIN
```

Query Match 90.8%; Score 526; DB 2; Length 1691;
Best Local Similarity 98.5%; Pred. No. 1e-156;
Matches 573; Conservative 0; Mismatches 5; Indels 4; Gaps 4;

```
Qy       1 ATCAGCTTGCCGTCGTCATAAGCAACTGGGCTTGCATTGCTTTTAGTTGTACAAACTGTG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1643 ATCAGCTTGCCGTCGTCATAAGCAACTGGGCTTGCATTGCTTTTAGTTGTACAAACTGTG 1584

Qy      61 CAGGCGTCTTCCAGCATTCTATTGTTCCGTGAAT-CCGGAAATCTGCACGTACCTGCTCC 119
           |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    1583 -AGGCGTCTTCCAGCATTCTATTGTTCCGTGAATCCCGGAAATCTGCACGTACCTGCTCC 1525

Qy     120 AGATTACTATGAGGATTATCCTTAAGTACAAGGGCCGCCGCCATCGTTCCGGTTCTTCCC 179
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1524 AGATTACTATGAGGATTATCCTTAAGTACAAGGGCCGCCGCCATCGTTCCGGTTCTTCCC 1465

Qy     180 ACTCCGCCCAGACAATGAATCATCGGTAAATGCTTATCTGATGAACTACGCCCCGGCGCG 239
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1464 ACTCCGCCCAGACAATGAATCATCGGTAAATGCTTATCTGATGAACTACGCCCCGGCGCG 1405

Qy     240 CCATTTTGGTTACTATTTTTCACCCTATCCGCCAGGTATTCTAACTGATCCGTAGACGGT 299
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1404 CCATTTTGGTTACTATTTTTCACCCTATCCGCCAGGTATTCTAACTGATCCGTAGACGGT 1345

Qy     300 AACGGCTGGTGATCTGGCCAATTTTTCACATGCAATACCGGGATTGTATACCGC-TTTCC 358
           |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db    1344 AACGGCTGGTGATCTGGCCAATTTTTCACATGCAATACCGGGATTGTATACCGCTTTTCC 1285

Qy     359 CCGCAGGACAGTTGCATATTGTATTGGTCTATCGCTTCTCCCTGACTGGCTGAGCTCTCT 418
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db    1284 CCGCAGGACAGTTGCATATTGTATTGGTCTATCGCTTCTCCCTGACTGGCTGAGCTCACT 1225

Qy     419 TTTTGGCTGTTGGTATGCACCTCGCCAAAGGTGTAGCTCCCTCTGAAATA-GGTGGTAAT 477
           ||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db    1224 TTTTGGCTGTTGGTATGCACCTCGCCAAAGGTGTAGCTCCCTCTGAAATAGGGTGGTAAT 1165

Qy     478 TGTTTTGCCTGCATCTGATCTTCCGACGTTAACACCACCAGGCACGAGCATTCTTTTTCA 537
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1164 TGTTTTGCCTGCATCTGATCTTCCGACGTTAACACCACCAGGCACGAGCATTCTTTTTCA 1105

Qy     538 AGAAGCATTTTCATATGCGCTTCCAGCGCATCCCGGCGATTT 579
```

```
           |||||||||||||||||||||||||||||||| |   ||||
Db    1104 AGAAGCATTTTCATATGCGCTTCCAGCGCATCCGGCGTATTT 1063
```

```
RESULT   2
STSIPSTP/c
LOCUS       STSIPSTP     3964 bp    DNA             BCT       28-JAN-1997
DEFINITION  S.typhi sipA, sipF, and stpA genes.
ACCESSION   X92546
NID         g1514650
VERSION     X92546.1  GI:1514650
KEYWORDS    sipA gene; sipF gene; stpA gene.
SOURCE      Salmonella typhi.
  ORGANISM  Salmonella typhi
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Salmonella.
REFERENCE   1  (bases 1 to 3964)
  AUTHORS   Arricau,N., Hermant,D., Waxin,H. and Popoff,M.Y.
  TITLE     Molecular characterization of the Salmonella typhi StpA protein
            that is related to both Yersinia YopE cytotoxin and YopH tyrosine
            phosphatase
  JOURNAL   Res. Microbiol. 148 (1), 21-26 (1997)
  MEDLINE   98068312
REFERENCE   2  (bases 1 to 3964)
  AUTHORS   Popoff,M.Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-OCT-1995) M.Y. Popoff, Institut Pasteur, Unite des
            Enterobacteries, 28 rue du Docteur Roux, 75724 Paris Cedex 15,
            FRANCE
  REMARK    Revised by submittor 09-APR-96
FEATURES             Location/Qualifiers
     source          1. .3964
                     /organism="Salmonella typhi"
                     /strain="Ty2"
                     /db_xref="taxon:601"
     gene            1. .1353
                     /gene="sipA"
     CDS             <1. .1353
                     /gene="sipA"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA63302.1"
                     /db_xref="PID:e208981"
                     /db_xref="PID:g1514651"
                     /db_xref="GI:1514651"
                     /db_xref="SPTREMBL:P74849"
                     /translation="SVGREARHLEMQTLIPLLLRNVFAQIPADKLPDPKIPEPAAGPV
                     PDGGKKAEPTGINININIDSSNHSVDNSKHINNSRSHVDNSQRHIDNSNHDNSRKTID
                     NSRTFIDNSQRHGESHHSTNSSNVSHSHSRVDSTTHQTETAHSASTGTIDHGIAGKID
                     VTAHATAEAVTNSSSSESKDGKVVTSEKGTTGETTSFDEVDGVTSKSIIGKPLQATVHG
                     VDDNKQQSQTAEIVNVKPLASQLAGVENVKIDTLQSDSTVITGNKAGTTDNDNSQTDK
                     TGPFSGLKFKQNSFLSTVPSVTNMHSIHFNAREAFLGVIRKALEPDASTPFPVRRAFD
                     GLRGEILPNDTIKSAALKAQCSDIDKHPELKAKMETLKEVITHHPQKEKLAEIALQFA
                     REAGLTRQKGETDYVLSNVLDGLIGDGSWRAGPAYESYLNKPGVDRVITTVDGLHMQR
                     "
     gene            1372. .1620
                     /gene="sipF"
     CDS             1372. .1620
                     /gene="sipF"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA63303.1"
                     /db_xref="PID:e208982"
                     /db_xref="PID:g1514652"
                     /db_xref="GI:1514652"
                     /db_xref="SPTREMBL:P74850"
                     /translation="MNMDIEARVKKVITSCIAVDVDSINGQTHLVEDLYPDSLDLIDI
                     VFGLSEEFDISCNENDLPDMTTFADICRVVKKSLESRV"
     gene            2330. .3961
                     /gene="stpA"
     CDS             2330. .3961
                     /gene="stpA"
                     /codon_start=1
```

```
                     /transl_table=11
                     /protein_id="CAA63304.1"
                     /db_xref="PID:e209029"
                     /db_xref="PID:g1514653"
                     /db_xref="GI:1514653"
                     /db_xref="SPTREMBL:P74851"
                     /translation="MLRYDERKLNNLTLSSFSKSGVSSDTRLYIAKENTDKAYVAPEK
                     FSSKVLTWLGKMPLFKNTEVVQKHTENIRVQNQKILQTFLQALTEKYGEKAVNNALYM
                     SSINMNKPLTQRLVVQITECVKGADGGFINLIKNKDNVGVMNAALVIKGGDTKVTEQN
                     NDVGAESKQPLLDIALKGLKRTIPQLEQMDGNSLRENFQEMASGNGPLRSLMTNLQSL
                     NKIPEAKQLNDYVTTLKNIQIGADRFSQWGTCGGEVERWIDKASTHELTQAVKKIHVI
                     AKELKNVTAELEKIKAGASMPQTMSGPTLGLARFAVSSIPINQQTQVKLSDGMPVPVN
                     TLTFDGKPVALAGSCPKNTPDALEAHMKMLLEKECSCLVVLTSEDQMQAKQLPAYFRG
                     SYTFGEVHTNSQKVSSASQGGAIDQYNMQLSCGEKRYTIPVLHVKNWPDHQPLPSTDQ
                     LEYLADRVKNSNQNGAPGRSSSDKHLPMIHCLGGVGRTGTMAAALVLKDNPHSNLEQV
                     RADFRNSRNNRMLEDASQFVQLKAMQAQLLMTTAS"
BASE COUNT     1264 a    759 c    966 g    975 t
ORIGIN
```

```
  Query Match             89.7%;  Score 519.6;  DB 1;  Length 3964;
  Best Local Similarity   97.8%;  Pred. No. 1.2e-154;
  Matches  569;  Conservative    0;  Mismatches    9;  Indels    4;  Gaps    4;

Qy       1 ATCAGCTTGCCGTCGTCATAAGCAACTGGGCTTGCATTGCTTTTAGTTGTACAAACTGTG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3962 ATCAGCTTGCCGTCGTCATAAGCAACTGGGCTTGCATTGCTTTTAGTTGTACAAACTGTG 3903

Qy      61 CAGGCGTCTTCCAGCATTCTATTGTTCCGTGAA-TCCGGAAATCTGCACGTACCTGCTCC 119
           |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    3902 -AGGCGTCTTCCAGCATTCTATTGTTCCGTGAATTCCGGAAATCTGCACGTACCTGCTCC 3844

Qy     120 AGATTACTATGAGGATTATCCTTAAGTACAAGGGCCGCCGCCATCGTTCCGGTTCTTCCC 179
           |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    3843 AGATTACTATGAGGATTATCCTTAAGTACAAGAGCCGCCGCCATCGTTCCGGTTCTTCCC 3784

Qy     180 ACTCCGCCCAGACAATGAATCATCGGTAAATGCTTATCTGATGAACTACGCCCCGGCGCG 239
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3783 ACTCCGCCCAGACAATGAATCATCGGTAAATGCTTATCTGATGAACTACGCCCCGGCGCG 3724

Qy     240 CCATTTTGGTTACTATTTTTCACCCTATCCGCCAGGTATTCTAACTGATCCGTAGACGGT 299
           |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db    3723 CCATTTTGGTTACTATTTTTCACCCTATCCGCCAGGTATTCTAACTGATCAGTAGACGGT 3664

Qy     300 AACGGCTGGTGATCTGGCCAATTTTTCACATGCAATACCGGGATTGTATACCGC-TTTCC 358
           |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db    3663 AACGGCTGGTGATCTGGCCAATTTTTCACATGCAATACCGGGATTGTATACCGCTTTTCC 3604

Qy     359 CCGCAGGACAGTTGCATATTGTATTGGTCTATCGCTTCTCCCTGACTGGCTGAGCTCTCT 418
           |||||||||||||||||||||||||||||||||||| |||||||||||||||||||| ||
Db    3603 CCGCAGGACAGTTGCATATTGTATTGGTCTATCGCTCCTCCCTGACTGGCTGAGCTCACT 3544

Qy     419 TTTTGGCTGTTGGTATGCACCTCGCCAAAGGTGTAGCTCCCTCTGAAATAGG-TGGTAAT 477
           ||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db    3543 TTTTGGCTGTTGGTATGCACCTCGCCAAAGGTGTAGCTCCCTCTGAAATAGGCTGGTAAT 3484

Qy     478 TGTTTTGCCTGCATCTGATCTTCCGACGTTAACACCACCAGGCACGAGCATTCTTTTTCA 537
           |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db    3483 TGTTTTGCCTGCATCTGATCTTCCGACGTTAACACCACCAGGCACGAACATTCTTTTTCA 3424

Qy     538 AGAAGCATTTTCATATGCGCTTCCAGCGCATCCCGGCGATTT 579
           |||||||||||||||||||||||||||||| |   ||||
Db    3423 AGAAGCATTTTCATATGCGCTTCCAGCGCATCCGGCGTATTT 3382
```

```
RESULT   3
STIAGAB
LOCUS       STIAGAB      2406 bp    DNA             BCT       28-JUN-1995
DEFINITION  S.typhi genes iagA and iagB.
ACCESSION   X80892
NID         g791139
VERSION     X80892.1  GI:791139
KEYWORDS    .
SOURCE      Salmonella typhi.
  ORGANISM  Salmonella typhi
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Salmonella.
REFERENCE   1  (bases 1 to 2406)
  AUTHORS   Miras,I., Hermant,D., Arricau,N. and Popoff,M.Y.
  TITLE     Nucleotide sequence of iagA and iagB genes involved in invasion of
            HeLa cells by Salmonella enterica subsp. enterica ser. Typhi
  JOURNAL   Res. Microbiol. 146 (1), 17-20 (1995)
  MEDLINE   95273801
REFERENCE   2  (bases 1 to 2406)
  AUTHORS   Popoff,M.Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-AUG-1994) M.Y. Popoff, Institut Pasteur, Unite des
            Enterobacteries, 28 rue du Docteur Roux, 75724 Paris Cedex 15,
            FRANCE
REFERENCE   3  (bases 1 to 2406)
  AUTHORS   Chevrier,D., Popoff,M.Y., Dion,M.P., Hermant,D. and Guesdon,J.L.
  TITLE     Rapid detection of Salmonella subspecies I by PCR combined with
            non-radioactive hybridisation using covalently immobilised
            oligonucleotide on a microplate
  JOURNAL   FEMS Immunol. Med. Microbiol. 10 (3-4), 245-501 (1995)
  MEDLINE   95291265
FEATURES             Location/Qualifiers
     source          1. .2406
                     /organism="Salmonella typhi"
                     /strain="Ty2"
                     /db_xref="taxon:601"
     gene            98. .1759
                     /gene="iagA"
     CDS             98. .1759
                     /gene="iagA"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA56853.1"
                     /db_xref="PID:g791140"
                     /db_xref="GI:791140"
                     /db_xref="SWISS-PROT:P43016"
                     /translation="MPHFNPVPVSNKKFVFDDFILNMDGSLVRSEKKVNIPPKEYAVL
                     VILLEAAGKIVSKNTLLDQVWGDAEVNEESLTRCIYALRRILSEDKEHRYIETLYGQG
                     YRFNRPVVVVSPPAPQPTTHTLAILPFQMQDQVQSESLHYSIVKGLSQYAPFGLSVLP
                     VTITKNCRSVKDILELMDQLRPDYYISGQMIPDGNDNIVQIEIVRVKGYHLLHQESIK
                     LIEHQPASLLQNKIANLLLRCIPGLRWDTKQISELNSIDSTMVYLRGKHELNQYTPYS
                     LQQALKLLTQCVNMSPNSIAPYCALAECYLSMAQMGIFDKQNAMIKAKEHAIKATELD
                     HNNPQALGLLGLINTIHSEYIVGSLLFKQANLLSPISADIKYYYGWNLFMAGQLEEAL
                     QTINECLKLDPTRAAAGITKLWITYYHTGIDDAIRLGDELRSQHLQDNPILLSMQVMF
                     LSLKGKHELARKLTKEISTQEITGLIAVNLLYAEYCQNSERALPTIREFLESEQRIDN
                     NPGLLPLVLVAHGEAIAEKMWNKFKNEDNIWFKRWKQDPRLIKLR"
     gene            1777. .2259
                     /gene="iagB"
     CDS             1777. .2259
                     /gene="iagB"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA56854.1"
                     /db_xref="PID:g791141"
                     /db_xref="GI:791141"
                     /db_xref="SWISS-PROT:P43018"
                     /translation="MHYFFIIVIWLLSINTAWADCWLQAEKMFNIESELLYAIAQQES
                     AMKPGAIGHNRDGSTDLGLMQINSFHMKRLKKMGISEKQLLQDPCISVIVGASILSDM
                     MKIYGFSWEAVGAYNAGTSPKRSDIRKRYAKKIWENYRKLKEMSAEEKNKRLSIAVNK
                     "
BASE COUNT      736 a    488 c    519 g    663 t
ORIGIN
```

```
  Query Match             14.2%;  Score 82.4;  DB 1;  Length 2406;
  Best Local Similarity   97.9%;  Pred. No. 1.2e-15;
  Matches   94;  Conservative    0;  Mismatches    1;  Indels    1;  Gaps    1;

Qy       1 ATCAGCTTGCCGTCGTCATAAGCAACTGGGCTTGCATTGCTTTTAGTTGTACAAACTGTG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2312 ATCAGCTTGCCGTCGTCATAAGCAACTGGGCTTGCATTGCTTTTAGTTGTACAAACTGTG 2371
```

```
Qy     61 CAGGCGTCTTCCAGCATTCTATTGTTCCGTGAATCC 96
          ||||||||||||||||||||||||||||||||||| |
Db   2372 -AGGCGTCTTCCAGCATTCTATTGTTCCGTGAATTC 2406


RESULT   4
A52315
LOCUS       A52315       2405 bp    DNA             PAT       12-DEC-1997
DEFINITION  Sequence 19 from Patent EP0721989.
ACCESSION   A52315
NID         g2851996
VERSION     A52315.1  GI:2851996
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 2405)
  AUTHORS   Popoff,M.Y. and Le,G.F.
  TITLE     Oligonucleotides for the detection of salmonella
  JOURNAL   Patent: EP 0721989-A 19 17-JUL-1996;
            PASTEUR INSTITUT (FR)
COMMENT     Other publication JP 8317798 961203
            Other publication CA 2167354 960717
            Other publication FR 2729392 960719.
FEATURES             Location/Qualifiers
     source          1. .2405
                     /organism="unidentified"
                     /db_xref="taxon:32644"
     CDS             97. .1758
                     /note="GENE IAGA"
                     /codon_start=1
                     /protein_id="CAA03394.1"
                     /db_xref="PID:e1215753"
                     /db_xref="PID:g2851997"
                     /db_xref="GI:2851997"
                     /translation="MPHFNPVPVSNKKFVFDDFILNMDGSLVRSEKKVNIPPKEYAVL
                     VILLEAAGKIVSKNTLLDQVWGDAEVNEESLTRCIYALRRILSEDKEHRYIETLYGQG
                     YRFNRPVVVVSPPAPQPTTHTLAILPFQMQDQVQSESLHYSIVKGLSQYAPFGLSVLP
                     VTITKNCRSVKDILELMDQLRPDYYISGQMIPDGNDNIVQIEIVRVKGYHLLHQESIK
                     LIEHQPASLLQNKIANLLLRCIPGLRWDTKQISELNSIDSTMVYLRGKHELNQYTPYS
                     LQQALKLLTQCVNMSPNSIAPYCALAECYLSMAQMGIFDKQNAMIKAKEHAIKATELD
                     HNNPQALGLLGLINTIHSEYIVGSLLFKQANLLSPISADIKYYYGWNLFMAGQLEEAL
                     QTINECLKLDPTRAAAGITKLWITYYHTGIDDAIRLGDELRSQHLQDNPILLSMQVMF
                     LSLKGKHELARKLTKEISTQEITGLIAVNLLYAEYCQNSERALPTIREFLESEQRIDN
                     NPGLLPLVLVAHGEAIAEKMWNKFKNEDNIWFKRWKQDPRLIKLR"
     CDS             1776. .2258
                     /note="SEQUENCE DU GENE IAGB"
                     /codon_start=1
                     /protein_id="CAA03395.1"
                     /db_xref="PID:e1215754"
                     /db_xref="PID:g2851998"
                     /db_xref="GI:2851998"
                     /translation="MHYFFIIVIWLLSINTAWADCWLQAEKMFNIESELLYAIAQQES
                     AMKPGAIGHNRDGSTDLGLMQINSFHMKRLKKMGISEKQLLQDPCISVIVGASILSDM
                     MKIYGFSWEAVGAYNAGTSPKRSDIRKRYAKKIWENYRKLKEMSAEEKNKRLSIAVNK
                     "
BASE COUNT      735 a    488 c    519 g    663 t
ORIGIN


  Query Match             14.2%;  Score 82.4;  DB 5;  Length 2405;
  Best Local Similarity   97.9%;  Pred. No. 1.2e-15;
  Matches   94;  Conservative    0;  Mismatches    1;  Indels    1;  Gaps    1;

Qy      1 ATCAGCTTGCCGTCGTCATAAGCAACTGGGCTTGCATTGCTTTTAGTTGTACAAACTGTG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2311 ATCAGCTTGCCGTCGTCATAAGCAACTGGGCTTGCATTGCTTTTAGTTGTACAAACTGTG 2370

Qy     61 CAGGCGTCTTCCAGCATTCTATTGTTCCGTGAATCC 96
          ||||||||||||||||||||||||||||||||||| |
Db   2371 -AGGCGTCTTCCAGCATTCTATTGTTCCGTGAATTC 2405


RESULT   5
I66494
LOCUS       I66494       7218 bp    DNA             PAT       23-DEC-1997
DEFINITION  Sequence 14 from patent US 5670367.
ACCESSION   I66494
NID         g2724471
VERSION     I66494.1  GI:2724471
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 7218)
  AUTHORS   Dorner,F., Scheiflinger,F. and Falkner,F.Gunter.
  TITLE     Recombinant fowlpox virus
  JOURNAL   Patent: US 5670367-A 14 23-SEP-1997;
FEATURES             Location/Qualifiers
     source          1. .7218
                     /organism="unknown"
BASE COUNT     1944 a   1491 c   1486 g   1929 t    368 others
ORIGIN


  Query Match             8.4%;  Score 48.6;  DB 5;  Length 7218;
  Best Local Similarity   4.5%;  Pred. No. 7.3e-05;
  Matches   18;  Conservative  216;  Mismatches  165;  Indels    0;  Gaps    0;

Qy    154 CCGCCGCCATCGTTCCGGTTCTTCCCACTCCGCCCAGACAATGAATCATCGGTAAATGCT 213
          :: :: :: :: :::: :::::::: :::: ::::   :  :   :: ::  :   : ::
Db   1072 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1131

Qy    214 TATCTGATGAACTACGCCCCGGCGCGCCATTTTGGTTACTATTTTTCACCCTATCCGCCA 273
          : :::  :   :: : ::::  : : :: :::: :: :: :::::: :::: ::: ::
Db   1132 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1191

Qy    274 GGTATTCTAACTGATCCGTAGACGGTAACGGCTGGTGATCTGGCCAATTTTTCACATGCA 333
           : ::::  ::  ::: :  :  :  :  ::  :  :::  ::  :::::: : : :
Db   1192 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1251

Qy    334 ATACCGGGATTGTATACCGCTTTCCCCGCAGGACAGTTGCATATTGTATTGGTCTATCGC 393
          : ::    :: : : :: :::::::: :    :  :: : : : : : ::  ::: :: :
Db   1252 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1311

Qy    394 TTCTCCCTGACTGGCTGAGCTCTCTTTTTGGCTGTTGGTATGCACCTCGCCAAAGGTGTA 453
          :::::::: ::  ::   :::::::::: :: ::  : : : :::: ::     : :
Db   1312 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1371

Qy    454 GCTCCCTCTGAAATAGGTGGTAATTGTTTTGCCTGCATCTGATCTTCCGACGTTAACACC 513
           :::::::    :  :  :  :: :::: ::: : :::  :::::  : ::  : ::
Db   1372 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1431

Qy    514 ACCAGGCACGAGCATTCTTTTTCAAGAAGCATTTTCATA 552
          ::   |   |   |||| | ||   ||  | || ||||
Db   1432 YYYGTACCAAATTCTTCTATCTCTTTAACTACTTGCATA 1470


RESULT   6
ATAC005561/c
LOCUS       ATAC005561  90859 bp    DNA             PLN       28-DEC-1998
DEFINITION  Arabidopsis thaliana chromosome II BAC T9F8 genomic sequence,
            complete sequence.
ACCESSION   AC005561
NID         g4063756
VERSION     AC005561.1  GI:4063756
KEYWORDS    HTG.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
            Arabidopsis.
REFERENCE   1  (bases 1 to 90859)
  AUTHORS   Rounsley,S.D., Lin,X., Kaul,S., Shea,T.P., Fujii,C.Y., Mason,T.M.,
```

```
            Shen,M., Ronning,C.M., Fraser,C.M., Somerville,C.R. and Venter,J.C.
  TITLE     Arabidopsis thaliana chromosome II BAC T9F8 genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 90859)
  AUTHORS   Rounsley,S.D. and Lin,X.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-AUG-1998) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, rounsley@tigr.org
REFERENCE   3  (bases 1 to 90859)
  AUTHORS   Lin,X.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-DEC-1998) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
COMMENT     On Dec 28, 1998 this sequence version replaced gi:3818357.
            Address all correspondence to:
            Xiaoying Lin
            The Institute for Genomic Research
            9712 Medical Center Dr.
            Rockville, MD 20850, USA
            e-mail: xlin@tigr.org
            BAC clone T9F8 is from Arabidopsis chromosome II and is near the
            molecular marker mi310.
            The orientation of the sequence is from SP6 to T7 end of the BAC
            clone.
            Genes were identified by a combination of three methods: Gene
            prediction programs including GRAIL (available by anonymous ftp
            from arthur.epm.ornl.gov), Genefinder (Phil Green, University of
            Washington), Genscan (Chris Burge,
            http://gnomic.stanford.edu/~chris/GENSCANW.html), and NetPlantGene
            (http://www.cbs.dtu.dk/netpgene/cbsnetpgene.html), searches of the
            complete sequence against a peptide database and the Arabidopsis
            EST database at TIGR (http://www.tigr.org/tdb/at/at.html).
            Annotated genes are named to indicate the level of evidence for
            their annotation. Genes with similarity to other proteins are named
            after the database hits. Genes without significant peptide
            similarity but with EST similarity are named as 'unknown' proteins.
            Genes without protein or EST similarity, that are predicted by more
            than two gene prediction programs over most of their length are
            annotated as 'hypothetical' proteins. Genes encoding tRNAs are
            predicted by tRNAscan-SE (Sean Eddy,
            http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are
            identified by repeatmasker (Arian Smit,
            http://ftp.genome.washington.edu/RM/RepeatMasker.html). Regions of
            genomic sequence that are not annotated as genes but have predicted
            exons by GRAIL are annotated as misc features.
FEATURES              Location/Qualifiers
     source           1. .90859
                      /organism="Arabidopsis thaliana"
                      /cultivar="Columbia"
                      /db_xref="taxon:3702"
                      /chromosome="II"
                      /map="mi310"
                      /clone="T9F8"
     misc_feature     23. .37
                      /note="exon predicted by xgrail, quality marginal"
     misc_feature     68. .321
                      /note="exon predicted by xgrail, quality excellent"
     misc_feature     complement(162. .317)
                      /note="exon predicted by xgrail, quality good_shadowexon"
     mRNA             complement(join(<541. .1296,1322. .>1369))
                      /gene="T9F8.1"
     gene             complement(<541. .>1369)
                      /gene="T9F8.1"
     CDS              complement(join(541. .1296,1322. .1369))
                      /gene="T9F8.1"
                      /codon_start=1
                      /product="putative retrotransposon"
                      /protein_id="AAC98468.1"
                      /db_xref="PID:g4063761"
                      /db_xref="GI:4063761"
                      /translation="MENRIPYHQKGKELRENRAPNLDTSDLVEENALTLIGRLTNQSG
                      QRLWALFLFLSNPWTLKGKAIGSDLGQGVFQLKFDFREDLQQVVDNRPYHFDQWMVIL
                      QKWEPIISPSFPSLISFWIELQGLPKHFWKPKMLQAIREELGEVMTKEITSSSVKIKV
                      LLDGLQPLIKETIVDFPNGGEAVVYLDYKNLKNHCLHCQRLSHEKKNCPRVQSAPVST
                      KTSNPLPVASLPVSQSNSRNYYTPADNFVAPKSPLTLWTENPPSPSGQL"
     misc_feature     1545. .1683
                      /note="exon predicted by xgrail, quality excellent"
     misc_feature     complement(2599. .2718)
                      /note="exon predicted by xgrail, quality marginal"
     repeat_region    2747. .2831
                      /rpt_family="(CAAAA)n"
     misc_feature     complement(2920. .3086)
                      /note="exon predicted by xgrail, quality marginal"
     misc_feature     2938. .3002
                      /note="exon predicted by xgrail, quality
                      marginal_shadowexon"
     repeat_region    3555. .3642
                      /rpt_family="(TAAAA)n"
     misc_feature     complement(3648. .3678)
                      /note="exon predicted by xgrail, quality good"
     misc_feature     complement(3728. .3785)
                      /note="exon predicted by xgrail, quality marginal"
     repeat_region    complement(3900. .4002)
                      /rpt_family="AT_rich"
     misc_feature     complement(4024. .4078)
                      /note="exon predicted by xgrail, quality good"
     repeat_region    4077. .4173
                      /rpt_family="(TAA)n"
     repeat_region    4185. .4220
                      /rpt_family="AT_rich"
     repeat_region    complement(4295. .4418)
                      /rpt_family="(TAAAA)n"
     repeat_region    complement(4990. .5026)
                      /rpt_family="AT_rich"
     misc_feature     5702. .5799
                      /note="exon predicted by xgrail, quality
                      excellent_shadowexon"
     repeat_region    complement(6099. .6139)
                      /rpt_family="AT_rich"
     repeat_region    complement(6187. .6253)
                      /rpt_family="AT_rich"
     misc_feature     complement(6836. .6908)
                      /note="exon predicted by xgrail, quality excellent"
     misc_feature     complement(6998. .7036)
                      /note="exon predicted by xgrail, quality marginal"
     misc_feature     complement(7349. .7374)
                      /note="exon predicted by xgrail, quality excellent"
     misc_feature     7444. .7473
                      /note="exon predicted by xgrail, quality
                      marginal_shadowexon"
     repeat_region    complement(8415. .8474)
                      /rpt_family="AT_rich"
     repeat_region    8709. .8767
                      /rpt_family="AT_rich"
     repeat_region    complement(9301. .9351)
                      /rpt_family="AT_rich"
     misc_feature     9401. .9485
                      /note="exon predicted by xgrail, quality good"
     misc_feature     complement(10059. .10178)
                      /note="exon predicted by xgrail, quality
                      marginal_shadowexon"
     misc_feature     10063. .10151
                      /note="exon predicted by xgrail, quality excellent"
     repeat_region    complement(10240. .10294)
                      /rpt_family="AT_rich"
     repeat_region    complement(10645. .10673)
                      /rpt_family="AT_rich"
     misc_feature     complement(11738. .11844)
                      /note="exon predicted by xgrail, quality good"
     misc_feature     complement(12020. .12052)
                      /note="exon predicted by xgrail, quality marginal"
     repeat_region    12425. .12513
                      /rpt_family="(TAAAA)n"
     repeat_region    12618. .12645
                      /rpt_family="AT_rich"
     repeat_region    complement(13593. .13614)
```

```
                     /rpt_family="AT_rich"
     gene            complement(13760. .17323)
                     /gene="T9F8.2"
                     /pseudo
     repeat_region   17291. .17378
                     /rpt_family="(GAAAA)n"
     repeat_region   complement(17590. .17623)
                     /rpt_family="AT_rich"
     misc_feature    18168. .18286
                     /note="exon predicted by xgrail, quality
                     marginal_shadowexon"
     misc_feature    complement(18174. .18236)
                     /note="exon predicted by xgrail, quality marginal"
     misc_feature    19756. .19848
                     /note="exon predicted by xgrail, quality
                     marginal_shadowexon"
     misc_feature    complement(20049. .20236)
                     /note="exon predicted by xgrail, quality good_shadowexon"
     misc_feature    complement(20268. .20364)
                     /note="exon predicted by xgrail, quality good_shadowexon"
     misc_feature    complement(21061. .21093)
                     /note="exon predicted by xgrail, quality
                     marginal_shadowexon"
     repeat_region   complement(21123. .21190)
                     /rpt_family="(TAAAA)n"
     repeat_region   complement(21137. .21191)
                     /rpt_family="POLY_A"
     misc_feature    complement(21207. .21498)
                     /note="exon predicted by xgrail, quality good_shadowexon"
     misc_feature    21219. .21324
                     /note="exon predicted by xgrail, quality excellent"
     misc_feature    21531. .21956
                     /note="exon predicted by xgrail, quality excellent"
     misc_feature    complement(21565. .21608)
                     /note="exon predicted by xgrail, quality good_shadowexon"
     misc_feature    complement(21660. .21775)
                     /note="exon predicted by xgrail, quality good_shadowexon"
     misc_feature    complement(21981. .22076)
                     /note="exon predicted by xgrail, quality good_shadowexon"

  Query Match             5.9%;  Score 34.2;  DB 8;  Length 90859;
  Best Local Similarity   47.4%;  Pred. No. 3.4;
  Matches  102;  Conservative    0;  Mismatches  113;  Indels    0;  Gaps    0;

Qy      13 TCGTCATAAGCAACTGGGCTTGCATTGCTTTTAGTTGTACAAACTGTGCAGGCGTCTTCC 72
           |||   |||  || ||  |||  | |  || |   ||||  || |||||
Db   33599 TCGGAGTAAAAAAGTGAACTTAAAATTTTTATTTCTGTAAGAATTGTGCACTTTCTCAAT 33540

Qy      73 AGCATTCTATTGTTCCGTGAATCCGGAAATCTGCACGTACCTGCTCCAGATTACTATGAG 132
           | ||||||||     | | ||||   || |    | | || ||  |      |||||
Db   33539 ATCATTCTATCAAAACTTCAATCTACAATTGCATAGGGACATGGCCGGTTGCACTATTTT 33480

Qy     133 GATTATCCTTAAGTACAAGGGCCGCCGCCATCGTTCCGGTTCTTCCCACTCCGCCCAGAC 192
             |   | ||  |   ||    | ||   |    | || |  ||     |   |
Db   33479 ACTATACATTTTGGGGAAAATGCACCTTTAAGCATTCGTTCTTTTTGTGTGGAGCTTATT 33420

Qy     193 AATGAATCATCGGTAAATGCTTATCTGATGAACTA 227
           || ||| |||  |   | ||| |||| |||| | |
Db   33419 AAAGAAGCATATGGGTAAGCTAATCTCATGATCAA 33385


RESULT   7
AC006415/c
LOCUS       AC006415   162346 bp    DNA             HTG       21-MAY-1999
DEFINITION  Drosophila melanogaster chromosome 2 clone BACR22A22 (D533) RPCI-98
            22.A.22 map 40B-40C strain y2; cn bw sp, WORKING DRAFT SEQUENCE, 1
            unordered pieces.
ACCESSION   AC006415
NID         g4883582
VERSION     AC006415.9  GI:4883582
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 162346)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Sequencing of Drosophila melanogaster
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 162346)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JAN-1999) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On May 21, 1999 this sequence version replaced gi:4699921.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu. All contigs in this submission meet
            the following cutoffs: length >= 400 bases, phrap computed error
            rate <= 1/10.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   162346: contig of 162346 bp in length.
FEATURES            Location/Qualifiers
     source         1. .162346
                    /organism="Drosophila melanogaster"
                    /strain="y2; cn bw sp"
                    /db_xref="taxon:7227"
                    /clone="BACR22A22 (D533) RPCI-98 22.A.22"
                    /chromosome="2"
                    /clone_lib="RPCI-98 (Roswell Park Cancer Institute
                    Drosophila melanogaster BAC library, partial EcoRI in
                    pBACe3.6"
                    /map="40B-40C"
BASE COUNT    48828 a  32586 c  32007 g  48925 t
ORIGIN


  Query Match             5.9%;  Score 34.2;  DB 35;  Length 162346;
  Best Local Similarity   53.3%;  Pred. No. 3.6;
  Matches   72;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy     305 CTGGTGATCTGGCCAATTTTTCACATGCAATACCGGGATTGTATACCGCTTTCCCCGCAG 364
           ||| |  ||| ||       |||||||| |||   |||  ||||    |  || |||   |
Db  126965 CTGCTATTCTAGCGTCAACTTCACATTCAAAGTCGGAGTTGTCCCTCTGTTGCCCATAAC 126906

Qy     365 GACAGTTGCATATTGTATTGGTCTATCGCTTCTCCCTGACTGGCTGAGCTCTCTTTTTGG 424
            |  |||| | | ||        ||||| ||| | | | ||| |||  | | ||    |
Db  126905 TATTTTTGCTTCTGGTTGGACATAATCGCGTCTGCATTATTGGTTGATGTTTGTTGCCTG 126846

Qy     425 CTGTTGGTATGCACC 439
           || | ||| | |
```

```
Db 126845 TACTTTGCATGGAGC 126831


RESULT   8
HSU90917/c
LOCUS       HSU90917     1318 bp    mRNA            PRI       30-MAR-1997
DEFINITION  Human clone 23641 mRNA sequence.
ACCESSION   U90917
NID         g1913898
VERSION     U90917.1  GI:1913898
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1318)
  AUTHORS   Andersson,B., Wentland,M.A., Ricafrente,J.Y., Liu,W. and Gibbs,R.A.
  TITLE     A 'double adaptor' method for improved shotgun library construction
  JOURNAL   Anal. Biochem. 236 (1), 107-113 (1996)
  MEDLINE   96207227
REFERENCE   2  (bases 1 to 1318)
  AUTHORS   Yu,W., Andersson,B., Worley,K.C., Muzny,D.M., Ding,Y., Liu,W.,
            Ricafrente,J.Y., Wentland,M.A., Lennon,G. and Gibbs,R.A.
  TITLE     Large-scale concatenation cDNA sequencing
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1318)
  AUTHORS   Yu,W. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-FEB-1997) Molecular and Human Genetics, Baylor
            College of Medicine, One Baylor Plaza S930, Houston, TX 77030, USA
COMMENT     similar to human putative M phase phosphoprotein 2 encoded by
            Genbank Accession Number L16783.
FEATURES             Location/Qualifiers
     source          1. .1318
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /sex="female"
                     /dev_stage="infant"
                     /tissue_type="brain"
                     /clone_lib="Soares library 1NIB from IMAGE consortium"
                     /clone="23641"
BASE COUNT      366 a    341 c    306 g    305 t
ORIGIN


  Query Match             5.8%;  Score 33.6;  DB 10;  Length 1318;
  Best Local Similarity   51.3%;  Pred. No. 3.7;
  Matches   78;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy     314 TGGCCAATTTTTCACATGCAATACCGGGATTGTATACCGCTTTCCCCGCAGGACAGTTGC 373
           | |||| |||| |   ||| |  || | |  |   ||| |||| |||    | |  | |
Db     228 TTGCCAGTTTTCCCAGTGCTAGTCCTGCACAGCTCACCTCTTTTCCCTGGCGTCTCTGAC 169

Qy     374 ATATTGTATTGGTCTATCGCTTCTCCCTGACTGGCTGAGCTCTCTTTTTGGCTGTTGGTA 433
             |   ||||    |     |   |||  |  || ||  || ||   || | ||  ||
Db     168 CTTGTTCCTTGGGACCTTTGGACCTGCTGGTTAACTTAGGCCTTTTGAGGGTTCTTAATA 109

Qy     434 TGCACCTCGCCAAAGGTGTAGCTCCCTCTGAA 465
           |     | |   ||| ||||  ||| |||||
Db     108 AGACAAGCTCACGAGGGGTAGGGCCCCCTGAA 77


RESULT   9
AC005911/c
LOCUS       AC005911   259037 bp    DNA             HTG       27-MAY-1999
DEFINITION  Homo sapiens chromosome 12p13.3 clone RPCI11-372B4, WORKING DRAFT
            SEQUENCE, 69 ordered pieces.
ACCESSION   AC005911
NID         g4895150
VERSION     AC005911.4  GI:4895150
KEYWORDS    HTG; HIGS_PHASE2.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 259037)
  AUTHORS   Muzny,D., Arenson,A.D., Brundage,E., Carvelli,K., Chen,E., Di,W.,
            Ding,Y., Dugan,S., Durbin,J., Forcum,J., Ganesh,R., Garcia,C.,
            Goodman,M., Gorrell,J.H., Haywood,M., Jackson,L., Kampal,R.,
            Karpathy,S., Leal,B., Li,Y., Liu,W., Logan,O., Lu,J., Ly,T.,
            Martinez,C., Oswal,G., Perez,L., Rashid,N.D., Rowland,K.,
            Savage,L., Scherer,S.S., Shen,H., Timms,K.M., Todd,J., Vo,Q.,
            Worley,K.C., Yu,W., Chinault,C., Nelson,D. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 259037)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-NOV-1998) Molecular and Human Genetics, Baylor
            College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 27, 1999 this sequence version replaced gi:4878040.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 69 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved
            *        1    43739: contig of 43739 bp in length
            *    43740    43767: gap of unknown length
            *    43768    87416: contig of 43649 bp in length
            *    87417    87444: gap of unknown length
            *    87445   120621: contig of 33177 bp in length
            *   120622   120649: gap of unknown length
            *   120650   144679: contig of 24030 bp in length
            *   144680   144706: gap of unknown length
            *   144707   165422: contig of 20716 bp in length
            *   165423   165449: gap of unknown length
            *   165450   181323: contig of 15874 bp in length
            *   181324   181350: gap of unknown length
            *   181351   190940: contig of 9590 bp in length
            *   190941   190967: gap of unknown length
            *   190968   196402: contig of 5435 bp in length
            *   196403   196429: gap of unknown length
            *   196430   198853: contig of 2424 bp in length
            *   198854   198880: gap of unknown length
            *   198881   200917: contig of 2037 bp in length
            *   200918   200944: gap of unknown length
            *   200945   202926: contig of 1982 bp in length
            *   202927   202953: gap of unknown length
            *   202954   204860: contig of 1907 bp in length
            *   204861   204887: gap of unknown length
            *   204888   206633: contig of 1746 bp in length
            *   206634   206660: gap of unknown length
            *   206661   208226: contig of 1566 bp in length
            *   208227   208253: gap of unknown length
            *   208254   209697: contig of 1444 bp in length
            *   209698   209724: gap of unknown length
            *   209725   211152: contig of 1428 bp in length
            *   211153   211179: gap of unknown length
            *   211180   212584: contig of 1405 bp in length
            *   212585   212611: gap of unknown length
            *   212612   213988: contig of 1377 bp in length
            *   213989   214015: gap of unknown length
            *   214016   215277: contig of 1262 bp in length
            *   215278   215304: gap of unknown length
            *   215305   216530: contig of 1226 bp in length
            *   216531   216557: gap of unknown length
            *   216558   217770: contig of 1213 bp in length
            *   217771   217797: gap of unknown length
            *   217798   219004: contig of 1207 bp in length
            *   219005   219031: gap of unknown length
            *   219032   220173: contig of 1142 bp in length
            *   220174   220200: gap of unknown length
```

```
*   220201   221338: contig of 1138 bp in length
*   221339   221365: gap of unknown length
*   221366   222497: contig of 1132 bp in length
*   222498   222524: gap of unknown length
*   222525   223654: contig of 1130 bp in length
*   223655   223681: gap of unknown length
*   223682   224809: contig of 1128 bp in length
*   224810   224836: gap of unknown length
*   224837   225955: contig of 1119 bp in length
*   225956   225982: gap of unknown length
*   225983   226945: contig of 963 bp in length
*   226946   226972: gap of unknown length
*   226973   227930: contig of 958 bp in length
*   227931   227957: gap of unknown length
*   227958   228895: contig of 938 bp in length
*   228896   228922: gap of unknown length
*   228923   229853: contig of 931 bp in length
*   229854   229880: gap of unknown length
*   229881   230786: contig of 906 bp in length
*   230787   230813: gap of unknown length
*   230814   231707: contig of 894 bp in length
*   231708   231734: gap of unknown length
*   231735   232625: contig of 891 bp in length
*   232626   232652: gap of unknown length
*   232653   233540: contig of 888 bp in length
*   233541   233567: gap of unknown length
*   233568   234447: contig of 880 bp in length
*   234448   234474: gap of unknown length
*   234475   235350: contig of 876 bp in length
*   235351   235377: gap of unknown length
*   235378   236244: contig of 867 bp in length
*   236245   236271: gap of unknown length
*   236272   237137: contig of 866 bp in length
*   237138   237164: gap of unknown length
*   237165   238028: contig of 864 bp in length
*   238029   238055: gap of unknown length
*   238056   238916: contig of 861 bp in length
*   238917   238943: gap of unknown length
*   238944   239796: contig of 853 bp in length
*   239797   239823: gap of unknown length
*   239824   240669: contig of 846 bp in length
*   240670   240696: gap of unknown length
*   240697   241535: contig of 839 bp in length
*   241536   241562: gap of unknown length
*   241563   242393: contig of 831 bp in length
*   242394   242420: gap of unknown length
*   242421   243248: contig of 828 bp in length
*   243249   243275: gap of unknown length
*   243276   243940: contig of 665 bp in length
*   243941   243967: gap of unknown length
*   243968   244729: contig of 762 bp in length
*   244730   244756: gap of unknown length
*   244757   245563: contig of 807 bp in length
*   245564   245590: gap of unknown length
*   245591   246387: contig of 797 bp in length
*   246388   246414: gap of unknown length
*   246415   247192: contig of 778 bp in length
*   247193   247219: gap of unknown length
*   247220   247992: contig of 773 bp in length
*   247993   248019: gap of unknown length
*   248020   248792: contig of 773 bp in length
*   248793   248819: gap of unknown length
*   248820   249478: contig of 659 bp in length
*   249479   249505: gap of unknown length
*   249506   250265: contig of 760 bp in length
*   250266   250292: gap of unknown length
*   250293   251043: contig of 751 bp in length
*   251044   251070: gap of unknown length
*   251071   251818: contig of 748 bp in length
*   251819   251845: gap of unknown length
*   251846   252590: contig of 745 bp in length
*   252591   252617: gap of unknown length
*   252618   253357: contig of 740 bp in length
*   253358   253384: gap of unknown length
*   253385   254074: contig of 690 bp in length
*   254075   254101: gap of unknown length
*   254102   254791: contig of 690 bp in length
*   254792   254818: gap of unknown length
*   254819   255330: contig of 512 bp in length
*   255331   255357: gap of unknown length
*   255358   256010: contig of 653 bp in length
*   256011   256037: gap of unknown length
*   256038   256669: contig of 632 bp in length
*   256670   256696: gap of unknown length
*   256697   257318: contig of 622 bp in length
*   257319   257345: gap of unknown length
*   257346   257958: contig of 613 bp in length
*   257959   257985: gap of unknown length
*   257986   258565: contig of 580 bp in length
*   258566   258592: gap of unknown length
*   258593   259037: contig of 445 bp in length.
FEATURES             Location/Qualifiers
     source          1. .259037
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RPCI11-372B4"
                     /chromosome="12p13.3"
BASE COUNT    65886 a  61561 c  61046 g  68542 t   2002 others
ORIGIN


  Query Match             5.8%;  Score 33.6;  DB 35;  Length 259037;
  Best Local Similarity   51.3%;  Pred. No. 5.8;
  Matches   78;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy     314 TGGCCAATTTTTCACATGCAATACCGGGATTGTATACCGCTTTCCCCGCAGGACAGTTGC 373
           | |||| |||| |   ||| |  || | |  |   ||| |||| |||    | |  |  |
Db  152969 TTGCCAGTTTTCCCAGTGCTAGTCCTGCACAGCTCACCTCTTTTCCCTGGCGTCTCTGAC 152910

Qy     374 ATATTGTATTGGTCTATCGCTTCTCCCTGACTGGCTGAGCTCTCTTTTTGGCTGTTGGTA 433
           |  |   ||||    |      |   |||  |  || ||  || ||   || | ||  ||
Db  152909 CTTGTTCCTTGGGACCTTTGGACCTGCTGGTTAACTTAGGCCTTTTGAGGGTTCTTAATA 152850

Qy     434 TGCACCTCGCCAAAGGTGTAGCTCCCTCTGAA 465
           |     | |   ||| ||||  ||| |||||
Db  152849 AGACAAGCTCACGAGGGGTAGGGCCCCCTGAA 152818


RESULT  10
AC005841
LOCUS       AC005841    97078 bp    DNA              PRI       01-MAY-1999
DEFINITION  Homo sapiens 12p13.3 PAC RPCI4-816N1 (Roswell Park Cancer Institute
            Human PAC Library) complete sequence.
ACCESSION   AC005841
NID         g4731044
VERSION     AC005841.3  GI:4731044
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 97078)
  AUTHORS   Muzny,D., Arenson,A.D., Bouck,J., Bunac,C., Chen,Z., Ding,Y.,
            Dugan,S., Durbin,J., Forcum,J., Garcia,C., Gorrell,J.H.,
            Gorrell,L.L., Hernandez,J., Issar,A., Jackson,L., Kneitz,S.,
            Kondejewski,N., Lau,S., Leal,B., Lee,E., Lichtarge,O., Liu,W.,
            Logan,O., Lu,J., Marondel,I., Martinez,C., Merscher,S., Miller,A.,
            Montgomery,K., Oswal,G., Pampell,L.R., Parish,B.J., Perez,L.,
            Rashid,N.D., Rives,C., Scherer,S.E., Shen,H., Shim,C., Simon,M.,
            Vo,Q., Williamson,A., Worley,K.C., Xhang,A.M., Yang,R., Yu,W.,
            Zhou,X., Kucherlapati,R., Nelson,D. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 97078)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
```

```
  JOURNAL   Submitted (22-OCT-1998) Molecular and Human Genetics, Baylor
            College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 97078)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAY-1999) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 1, 1999 this sequence version replaced gi:4454408.
            INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
            gc-help@bcm.tmc.edu

            CLONE LENGTH: This sequence does not necessarily represent the
            entire insert of this clone. Overlapping regions of clones are only
            sequenced and submitted once, so the sequence for the remainder of
            the insert may be found in the record for the adjacent clones.
            Overlapping clones are noted at the beginning and end of the
            Features listing.

            ANNOTATION OF FEATURES:
               STSs are identified using ePCR (Genome Res. 7:541-550) searches
            of a local database that includes entries from dbSTS, GDB, and
            local mapping efforts.
               Repeats are identified using RepeatMasker (A. Smit and P. Green,
            unpublished.) for Human and Mouse sequences.
               Genes and Region of sequence similarity are identified by BLAST
            (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
            EST and cDNA sequences. Genes demonstrate at least two exons
            flanked by consensus splice sites that maintained sequence
            continuity across the splice junctions. Sequences that are not
            identical matches are annotated as similar.

            SEQUENCING READ COVERAGE:Sequencing is completed to a minimum
            standard of double strand coverage with a minimum of 2 clones and 2
            reads with no ambiguities or 2 chemistries with a minimum of 2
            clones and 3 reads with no ambiguities. If the sequence quality for
            a region does not meet this standard, it will be indicated in the
            annotation as Low Coverage.

            QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality
            standards - estimated error rate less than 1 per 10,000 bases.
            Reports of lowest quality individual bases and measures of base
            quality are listed below. Description of the metrics can be found
            at URL: http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.
            html.

            QUALSTAT-REPORT.
FEATURES             Location/Qualifiers
     source          1. .97078
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RPCI4-816N1"
                     /chromosome="12p13.3"
     repeat_region   1. .252
                     /rpt_family="AluSq"
     repeat_region   494. .562
                     /rpt_family="MER91C"
     repeat_region   712. .1021
                     /rpt_family="AluSx"
     misc_feature    1011. .1179
                     /function="Low coverage"
     gene            join(1477. .2027,3757. .3908,7250. .7441,9483. .9611,
                     11252. .11321,11509. .11684,16341. .18321)
                     /gene="Human hepatocyte nuclear factor-3/fork head homolog
                     11B (HFH-11B) mRNA U74613"
     repeat_region   2127. .2434
                     /rpt_family="AluSx"
     repeat_region   complement(2459. .2572)
                     /rpt_family="FLAM_C"
     misc_feature    2564. .2654
                     /function="Low coverage"
     repeat_region   complement(2577. .2888)
                     /rpt_family="AluSx"
     repeat_region   complement(2941. .3149)
                     /rpt_family="AluJo"
     repeat_region   complement(3376. .3676)
                     /rpt_family="AluSx"
     repeat_region   4034. .4186
                     /rpt_family="L2"
     repeat_region   complement(4341. .4638)
                     /rpt_family="AluJb"
     repeat_region   complement(4792. .5021)
                     /rpt_family="L1PA15"
     repeat_region   complement(5022. .5325)
                     /rpt_family="AluSg"
     repeat_region   complement(5328. .5621)
                     /rpt_family="AluSc"
     repeat_region   complement(5622. .5704)
                     /rpt_family="L1PA15"
     repeat_region   5729. .6018
                     /rpt_family="AluSp"
     repeat_region   6020. .6108
                     /rpt_family="L1ME3"
     repeat_region   6118. .6268
                     /rpt_family="AluJb"
     repeat_region   6269. .6567
                     /rpt_family="AluSq"
     repeat_region   6568. .6733
                     /rpt_family="AluJb"
     repeat_region   complement(8001. .8310)
                     /rpt_family="AluSx"
     repeat_region   complement(8329. .8493)
                     /rpt_family="AluY"
     repeat_region   complement(8543. .8821)
                     /rpt_family="AluJb"
     repeat_region   complement(8834. .9133)
                     /rpt_family="AluSx"
     repeat_region   complement(9860. .10008)
                     /rpt_family="MER5A"
     repeat_region   9891. .10026
                     /rpt_family="MER5A"
     repeat_region   10053. .10360
                     /rpt_family="AluSp"
     repeat_region   12000. .12299
                     /rpt_family="AluSg"
     repeat_region   12310. .12614
                     /rpt_family="AluJo"
     repeat_region   complement(12728. .12943)
                     /rpt_family="MER20"
     repeat_region   13136. .13162
                     /rpt_family="AT_rich"
     repeat_region   13163. .13462
                     /rpt_family="AluJo"
     repeat_region   13463. .13493
                     /rpt_family="AT_rich"
     repeat_region   complement(13806. .14105)
                     /rpt_family="AluSg"
     repeat_region   complement(14114. .14550)
                     /rpt_family="L1ME3A"
     repeat_region   14555. .14578
                     /rpt_family="AT_rich"
     repeat_region   complement(14579. .14672)
                     /rpt_family="L1PB1"
     repeat_region   complement(14822. .15117)
                     /rpt_family="AluSg"
     repeat_region   complement(15126. .15412)
                     /rpt_family="AluSx"
     repeat_region   complement(15493. .15765)
                     /rpt_family="AluJb"
     repeat_region   complement(15974. .16055)
                     /rpt_family="HY4"
     gene            complement(join(16213. .16470,17673. .18716))
                     /gene="Human clone 23641 mRNA sequence U90917"
     STS             16241. .16368
                     /gene="Human clone 23641 mRNA sequence U90917"
                     /standard_name="G21691"
```

```
                     /db_xref="dbSTS:35529"
     STS             17458. .17568
                     /gene="Human hepatocyte nuclear factor-3/fork head homolog
                     11B (HFH-11B) mRNA U74613"
                     /standard_name="G07228"
                     /db_xref="dbSTS:12725"
     repeat_region   19342. .19633
                     /rpt_family="AluSq"
     repeat_region   19730. .19920
                     /rpt_family="MIR"
     repeat_region   complement(21302. .21394)
                     /rpt_family="L2"
     repeat_region   21547. .21664
                     /rpt_family="MIR"
     repeat_region   21680. .21973
                     /rpt_family="AluSp"
     repeat_region   22382. .22426
                     /rpt_family="(CAAAAA)n"
     repeat_region   22432. .22728
                     /rpt_family="AluSq"
     repeat_region   complement(22896. .22978)
                     /rpt_family="MIR"
     repeat_region   complement(22979. .23281)
                     /rpt_family="AluSx"
     misc_feature    23214. .23274
                     /function="Low coverage"
     repeat_region   complement(23282. .23339)
                     /rpt_family="MIR"

  Query Match              5.8%;  Score 33.6;  DB 42;  Length 97078;
  Best Local Similarity   51.3%;  Pred. No. 5.3;
  Matches   78;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy     314 TGGCCAATTTTTCACATGCAATACCGGGATTGTATACCGCTTTCCCCGCAGGACAGTTGC 373
           | |||| |||| |   ||| |  || | |  |   ||| |||| |||    | |  | |
Db   18489 TTGCCAGTTTTCCCAGTGCTAGTCCTGCACAGCTCACCTCTTTTCCCTGGCGTCTCTGAC 18548

Qy     374 ATATTGTATTGGTCTATCGCTTCTCCCTGACTGGCTGAGCTCTCTTTTTGGCTGTTGGTA 433
             |  |   ||||    |     |   |||  |  || ||  || ||   || | ||  ||
Db   18549 CTTGTTCCTTGGGACCTTTGGACCTGCTGGTTAACTTAGGCCTTTTGAGGGTTCTTAATA 18608

Qy     434 TGCACCTCGCCAAAGGTGTAGCTCCCTCTGAA 465
             |     | |   ||| |||| ||| |||||
Db   18609 AGACAAGCTCACGAGGGGTAGGGCCCCCTGAA 18640


RESULT  11
CEW07E11
LOCUS       CEW07E11    14200 bp    DNA             INV       23-NOV-1998
DEFINITION  Caenorhabditis elegans cosmid W07E11, complete sequence.
ACCESSION   Z49868
NID         g870977
VERSION     Z49868.1  GI:870977
KEYWORDS    HTG; FMRFamide neuropeptide; Glutamate synthase.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 14200)
  AUTHORS   Mortimore,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUN-1995) Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
REFERENCE   2  (bases 1 to 14200)
  AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
  TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
  JOURNAL   Nature 368 (6466), 32-38 (1994)
  MEDLINE   94150718
COMMENT     Coding sequences below are predicted from computer analysis, using
            predictions from Genefinder (P. Green, U. Washington), and other
            available information.
            For a graphical representation of this sequence and its analysis
            see:-
            http://webace.sanger.ac.uk/cgi-
            bin/display?db=wormace&class=Sequence &object=W07E11
            Current sequence finishing criteria for the C. elegans genome
            sequencing consortium are that all bases are either sequenced
            unambiguously on both strands, or on a single strand with both a
            dye primer and dye terminator reaction, from distinct subclones.
            Exceptions are indicated by an explicit note.
            IMPORTANT:  This sequence is NOT necessarily the entire insert of
            the specified clone.  It may be shorter because we only sequence
            overlapping sections once, or longer because we arrange for a small
            overlap between neighbouring submissions.
            IMPORTANT: This sequence is not the entire insert of clone W07E11.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we arrange for a small overlap between
            neighbouring submissions.
            The true left end of clone W07E11 is at 1 in this sequence. The
            true right end of clone W07E11 is at 20366 in
            sequence Z49889.
            The true left end of clone T06H11 is at 14102 in this sequence. The
            true right end of clone ZC373 is at 8091 in this sequence. The
            start of this sequence (1. .103) overlaps with the end of sequence
            Z49131.
            The end of this sequence (14102. .14200) overlaps with the start of
            sequence Z49889.
FEATURES                Location/Qualifiers
     source             1. .14200
                        /organism="Caenorhabditis elegans"
                        /db_xref="taxon:6239"
                        /chromosome="X"
                        /clone="W07E11"
     CDS                join(2574. .2630,3196. .3494,3748. .4369,4691. .4877,
                        5483. .5637,14102. .14200,Z49889.1:100. .156,
                        Z49889.1:215. .487,Z49889.1:834. .986,Z49889.1:1170. .1248,
                        Z49889.1:1858. .2279,Z49889.1:2398. .2560,
                        Z49889.1:3533. .3627,Z49889.1:3680. .4237,
                        Z49889.1:4367. .5628,Z49889.1:5674. .6061,
                        Z49889.1:6310. .6465,Z49889.1:6669. .7300,
                        Z49889.1:7348. .8146,Z49889.1:8194. .8361)
                        /gene="W07E11.1"
                        /note="similar to glutamate synthase; cDNA EST EMBL:D27720
                        comes from this gene; cDNA EST EMBL:D27719 comes from this
                        gene; cDNA EST EMBL:D32507 comes from this gene; cDNA EST
                        EMBL:D33013 comes from this gene; cDNA EST EMBL:D33040
                        comes from this gene; cDNA EST EMBL:D35161 comes from this
                        gene; cDNA EST EMBL:D35739 comes from this gene; cDNA EST
                        EMBL:D35772 comes from this gene; cDNA EST EMBL:D37577
                        comes from this gene; cDNA EST yk378b5.3 comes from this
                        gene; cDNA EST yk378b5.5 comes from this gene; cDNA EST
                        yk405g6.3 comes from this gene; cDNA EST yk405g6.5 comes
                        from this gene; cDNA EST yk453g9.3 comes from this gene;
                        cDNA EST yk453g9.5 comes from this gene; cDNA EST
                        yk452h6.3 comes from this gene; cDNA EST yk452h6.5 comes
                        from this gene; cDNA EST yk420b11.3 comes from this gene;
                        cDNA EST yk420b11.5 comes from this gene; cDNA EST
                        yk355a5.3 comes from this gene; cDNA EST yk355a5.5 comes
                        from this gene; cDNA EST yk310b3.3 comes from this gene;
                        cDNA EST yk310b3.5 comes from this gene; cDNA EST
                        yk305c7.3 comes from this gene; cDNA EST yk305c7.5 comes
                        from this gene; cDNA EST yk275h4.3 comes from this gene;
                        cDNA EST yk275h4.5 comes from this gene; cDNA EST
                        yk321d5.3 comes from this gene; cDNA EST yk321d5.5 comes
                        from this gene; cDNA EST yk270d1.3 comes from this gene;
```

cDNA EST yk270d1.5 comes from this gene; cDNA EST yk247c12.3 comes from this gene; cDNA EST yk247c12.5 comes from this gene; cDNA EST yk248e10.3 comes from this gene; cDNA EST yk248e10.5 comes from this gene; cDNA EST yk246d12.3 comes from this gene; cDNA EST yk230h6.3 comes from this gene; cDNA EST yk230h6.5 comes from this gene; cDNA EST yk317a2.5 comes from this gene; cDNA EST yk349h8.5 comes from this gene; cDNA EST yk353c1.5 comes from this gene; cDNA EST yk359c1.5 comes from this gene"
/codon_start=1
/protein_id="CAA90032.1"
/db_xref="PID:e1350136"
/db_xref="PID:g3880576"
/db_xref="GI:3880576"
/translation="MCCFFRHRESMPIAPNYVPVYCILKQKSHGLIMVILTKEQQQEA
ANTGLWLPQLERDACGVGFVCSIKGTTSSKIMSDARTMLERMAHRGACGCDNDSGDGA
GVLTAIPDDLYRKSVKEQDGSDLPPLGQYATGILFLEAESYKQAKEAFQDLARACGLR
VVAWRKLGTNRECIGEEAKKTEPLIRQVFVSADYAESDPAKFERSVYLLRKQAVNSMT
KQEIECYVCSLSTSTIVYKGQFNTHQLFKFYDDLTNPEYQTHLALVHSRFSTNTFPSW
NRAQPNRILAHNGEINTLRGNINLMRAREGVMKSKYYRDDLQKLFPIVEEGLTDSGCL
DNVMEFLVRAGGRSLPEAAMTMVPEAWEKDDDMSTEKKHFYRWAAMSMEPWDGPALLA
FSDGRYIGAILDRNGLRPARYYLTDDDHLYLSSEVGVNDIPIEKVVKKDRLKPGRMLL
VDTHLKKIELDEDLKTRIALSRPHKQLSASRIYLDLIRKDDVLSHGAVTNEYLNRLHL
EQVNPGLMKKKDVHLDSDRRLALYAYTHDTFSLLLVPMIKEKKEALGSMGNDAALACL
SDYSPLLFSYFQQLFAQVTNPPIDPFREQIVMSLRCPIGPESNMLEPDTELASRLILE
QPVLSMVDMEVIKRTMYKGWKAKEIDITYPVKYGVKGLVPGLDAMCCVACTAALDGYQ
ILVLSDRNASSERVPIPSLLAVGAIHQCLIRHRLRMKVAIVVETGEARVVHDFCVLLG
FGADAICPYMVYETMHRLRNLGLLDKELNDDQVYQGYRQGVERGIFKVMAKMGISTLH
SYKHAQIFEIVGLAKDVVDMCFKNTVSRLGGATFEILAAEALKRHRSAFPTTSDASFG
DSKTLVASGTFHWRAGGEKHINEPLAIAKLQAAARLNNSKTFQEYSLASNMAQRWCTL
RGQLEIKTSKKIQIPLSEVEPASEIVKKFVTGAMSFGSISWETHTALAIAMNRIGGKS
NTGEGGEKPERYRKDQDPNQNLRSAIKQVASARFGVTSSYLANADELQIKMAQGAKPG
EGGELPGHKVTQDIADTRKSTAGVGLISPPPHHDIYSIEDLAQLIYDLKCANPVARVS
VKLVSEAGVGIVAAGVAKGNADHITVSGHDGGTGASSWTGIKHAGLPWELGVAETHQV
LTMNNLRSRVVLQADGQIRTGRDVMIAALLGADEFGMSTAPLIVLGCTMMRKCHLNTC
PVGVATQDPVLRAKFDGKPEHVVNYMFMVAEEVRYFLSKLGLRKLEDAVGRTDLLYAS
SNPVNKKATMLEFGSILKNAQQMFPNISIKGGSVKQVIELGALETRMLTELDGVFNDA
GEHKVFENEYITNIDRTFGTRISYEISKRYGELGLEGSRSITIGLKGHAGQSFCAFLA
KGVSVTLEGDANDYVGKCLSGGKIVIFPPKAASYKSEENSVIGNVALYGATSGDCWFR
GVAGERFAVRNSGANVIVEAVGDHGCEYMTGGRVIVLGSIGRNFAAAMSGGIAYLFAQ
EDNFSRLINAATVDLDDASTEDLIFVKAKIEEFVKLTGSELGQRILTNWQKEHQKIIK
VFPRDYKRVLAEQEEARKKQAELDKANQDMSLLSVDDTKPRTRTLSMDMQIAPTIRPH
NKKQKTALRKTSLLGDKRKQRRLSKSLRPQDMAGFENEEDLQDAADMEEEEESKSDLE
DHSDEEEFVEKKASVDIESLGIPAHLRKKDEPLDKLRGFVKYNRQKKMYRDPKERLND
WDEVYDFEAVRSNIREQAARCMDCGVPFCQGHSGCPLGNIIPKWNDFVFKKNWRQALE
QLLQTNNFPEFTGRVCPAPCEGACTLGIGSPAVTIKSIECAIIDYAFMQGWMKPCKPA
FNTGKRIAIIGSGPSGLGAAAQLVKVGHTVVVYERKNRVGGLLRYGIPTMKLDKFVVD
RRVSLLEQEGVRFLTNTEIGKHVPADFLLKENDAIIVCTGSTTARDLNVEGRDAKGVC
FAMEYLEKSQRRRAGDDVSWEGLDPANKKVIILGGGDTATDCIATSNRLGCKTVGAFE
ILPQPGPNRKPENPWPEWPLIFRVDYGHEEAKEMTGSDPRTYSVSTKRFLTTTNAAGI
TVLTGLEIVDVEWEKDDKGAWKLIEKNETIRTIECDLCILAMGFVGPEKSVIEQLNLK
TDPRSNILTPKDKYDSDVAKVFAAGDCRRGQSLVVWAIHEGRQAARQVDEYLMGKTTL
AGPGGIVTATIQLKNAI"
gene complement(10787. .11486)
/gene="W07E11.2"
CDS complement(join(10787. .10902,10951. .11127,11177. .11374, 11423. .11486))
/gene="W07E11.2"
/note="similar to FMRFamide neuropeptide repeats"
/codon_start=1
/protein_id="CAA90030.1"
/db_xref="PID:e1350134"
/db_xref="PID:g3880574"
/db_xref="GI:3880574"
/db_xref="SPTREMBL:Q23212"
/translation="MISPNHLILLFCVNCAFLVASDATPKRSPLGTMRFGKRAIADEM
TFEEDGYYPSNVMWKRSTVDSSEPVIRDQRTPLGTMRFGKRSAEPFGTMRFGKRNPEN
DTPFGTMRFGKRASEDALFGTMRFGKREDGNAPFGTMKFGKREAEEPLGTMRFGKRSA
DDSAPFGTMRFGKRNPLGTMRFGK"
gene complement(13197. .13723)
/gene="W07E11.3"
CDS complement(join(13197. .13242,13449. .13723))
/gene="W07E11.3"
/codon_start=1
/protein_id="CAA90031.1"
/db_xref="PID:e1350135"
/db_xref="PID:g3880575"
/db_xref="GI:3880575"
/db_xref="SPTREMBL:Q23213"
/translation="MQVSGILSALFLVLLAVIVSPFQFVQPKRILPIPTSRDQLLRGQ
LAYLKGTTVAQPAVNDNTLGIFEASAMAKRLRGEPIRFGKRSPREPIRFGKRFNPLPD
YDFQ"
BASE COUNT 4380 a 2458 c 2839 g 4523 t
ORIGIN

Query Match 5.8%; Score 33.4; DB 36; Length 14200;
Best Local Similarity 49.2%; Pred. No. 5.3;
Matches 88; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

Qy 389 ATCGCTTCTCCCTGACTGGCTGAGCTCTCTTTTTGGCTGTTGGTATGCACCTCGCCAAAG 448
Db 3656 ATTACTTGTTTCAGGCAGAAAATTCCCTGATATTCACTCTTATATTTCTCTAGCTGGCAT 3715

Qy 449 GTGTAGCTCCCTCTGAAATAGGTGGTAATTGTTTTGCCTGCATCTGATCTTCCGACGTTA 508
Db 3716 CTTTACGTTTTTCTCAAATAGTTATATTTCAGGCAGGATGGATCTGATCTCCCACCATTG 3775

Qy 509 ACACCACCAGGCACGAGCATTCTTTTTCAAGAAGCATTTTCATATGCGCTTCCAGCGCA 567
Db 3776 GGTCAATACGCCACGGGAATTCTATTTCTTGAAGCTGAATCGTACAAGCAGGCAAAGGA 3834

RESULT 12
AC005472/c
LOCUS AC005472 84981 bp DNA HTG 24-MAY-1999
DEFINITION Drosophila melanogaster chromosome 2 clone DS07138 (D387) map 49D1-49D3 strain y2; cn bw sp, WORKING DRAFT SEQUENCE, 6 unordered pieces.
ACCESSION AC005472
NID g4885688
VERSION AC005472.9 GI:4885688
KEYWORDS HTG; HTGS_PHASE1.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 84981)
AUTHORS Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L., Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E., Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and Rubin,G.M.
TITLE Sequencing of Drosophila melanogaster
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 84981)
AUTHORS Celniker,S.E., George,R.A., Galle,R., Svirskas,R.R., Hoskins,R.A., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Chavez,C., Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B., Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,E., Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R., Zieran,L.L. and Kimmel,B.E.
TITLE Direct Submission
JOURNAL Submitted (14-AUG-1998) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT On May 24, 1999 this sequence version replaced gi:4883584.
For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (http://www.fruitfly.org/sequence/) or send email

```
            to bdgp@fruitfly.berkeley.edu. All contigs in this submission meet
            the following cutoffs: length >= 400 bases, phrap computed error
            rate <= 1/10. P1 library location: 75-34.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      489: contig of 489 bp in length
            *      490      569: gap of unknown length
            *      570     1009: contig of 440 bp in length
            *     1010     1089: gap of unknown length
            *     1090     1548: contig of 459 bp in length
            *     1549     1628: gap of unknown length
            *     1629     2172: contig of 544 bp in length
            *     2173     2252: gap of unknown length
            *     2253     8390: contig of 6138 bp in length
            *     8391     8470: gap of unknown length
            *     8471    84981: contig of 76511 bp in length.
FEATURES             Location/Qualifiers
     source          1. .84981
                     /organism="Drosophila melanogaster"
                     /strain="y2; cn bw sp"
                     /db_xref="taxon:7227"
                     /clone="DS07138 (D387)"
                     /chromosome="2"
                     /clone_lib="P1 library, partial Sau3A in pNS582tet14Ad10"
                     /map="49D1-49D3"
BASE COUNT    25196 a  18504 c  17821 g  23060 t    400 others
ORIGIN


  Query Match             5.7%;  Score 33;  DB 35;  Length 84981;
  Best Local Similarity   58.8%;  Pred. No. 8.2;
  Matches   57;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy     333 AATACCGGGATTGTATACCGCTTTCCCCGCAGGACAGTTGCATATTGTATTGGTCTATCG 392
           ||| || ||| || | ||  |  ||  |  |  | | || ||||||||||     | ||||
Db   46999 AATTCCTGGAATGAAAACTCCGCTCGGCTAATTATATTTACATATTGTAAAACACCATCG 46940

Qy     393 CTTCTCCCTGACTGGCTGAGCTCTCTTTTTGGCTGTT 429
           | | |   | |   |||    |  |||||| |||||
Db   46939 CGTTTGTATCAGCAGCTATAATTGCTTTTTGTCTGTT 46903


RESULT  13
S72379/c
LOCUS       S72379       3816 bp                    INV       10-JUL-1992
DEFINITION  vg=vestigial [Drosophila melanogaster, Genomic/mRNA, 3816 nt].
ACCESSION   S72379
NID         g240993
VERSION     S72379.1  GI:240993
KEYWORDS    .
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 3816)
  AUTHORS   Williams,J.A., Bell,J.B. and Carroll,S.B.
  TITLE     Control of Drosophila wing and haltere development by the nuclear
            vestigial gene product
  JOURNAL   Genes Dev. 5 (12B), 2481-2495 (1991)
  MEDLINE   92090727
  REMARK    GenBank staff at the National Library of Medicine created this
            entry [NCBI gibbsq 72379] from the original journal article.
            This sequence comes from Figure 4.
FEATURES             Location/Qualifiers
     source          1. .3816
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
     gene            1. .3816
                     /partial
                     /gene="vg"
                     /note="vestigial"
                     /allele=""
                     /db_xref="FlyBase:FBgn0003975"
     CDS             1800. .3161
                     /gene="vg"
                     /note="vestigial; This sequence comes from Figure 4"
                     /codon_start=1
                     /protein_id="AAB20671.1"
                     /db_xref="PID:g240994"
                     /db_xref="GI:240994"
                     /translation="MAVSCPEVMYGAYYPYLYGRAGTSRSFYQYERFNQDLYSSSGVN
                     LAASSSASGSSHSPCSPILPPSVSANAAAAVAAAAHNSAAAAVAVAANQASSSGGIGG
                     GGLGGLGGLGGGPASGLLGSNVVPGSSSVGSVGLGMSPVLSGAAGHSLHSSHRTHAHS
                     LAHAHTHPHSHTHTHTHQTKEEDLIVPRSEAEARLVGSQQHQHHNESSCSSGPDSPRH
                     AHSHSHPLHGGGGATGGPSSAGGTGSGGGDGGGTGAIPKNLPALETPMGSGGGGLAGS
                     GQGQAQYLSASCVVFTNYSGDTASQVDEHFSRALNYNNKDSKESSSPMSNRNFPPSFW
                     NSNYVHPIPAPTHHQVSDLYGTATDTGYATDPWVPHAAHYGSYAHAAHAHAAHAHAYH
                     HNMAQYGSLLRLPQQYASHGSRLHHDQQTAHALEYSSYPTMAGLEAQVAQVQESSKDL
                     YWF"
BASE COUNT     1114 a    982 c    878 g    842 t
ORIGIN


  Query Match             5.7%;  Score 33;  DB 36;  Length 3816;
  Best Local Similarity   58.8%;  Pred. No. 6.3;
  Matches   57;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy     333 AATACCGGGATTGTATACCGCTTTCCCCGCAGGACAGTTGCATATTGTATTGGTCTATCG 392
           ||| || ||| || | ||  |  ||  |  |  | | || ||||||||||     | ||||
Db    1410 AATTCCTGGAATGAAAACTCCGCTCGGCTAATTATATTTACATATTGTAAAACACCATCG 1351

Qy     393 CTTCTCCCTGACTGGCTGAGCTCTCTTTTTGGCTGTT 429
           | | |   | |   |||    |  |||||| |||||
Db    1350 CGTTTGTATCAGCAGCTATAATTGCTTTTTGTCTGTT 1314


RESULT  14
DRU34590
LOCUS       DRU34590     1369 bp    mRNA            VRT       28-SEP-1995
DEFINITION  Danio rerio LIM homeobox protein (lim3) mRNA, complete cds.
ACCESSION   U34590
NID         g1000341
VERSION     U34590.1  GI:1000341
KEYWORDS    .
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
            Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;
            Cyprinoidea; Cyprinidae; Rasborinae; Danio.
REFERENCE   1  (bases 225 to 782)
  AUTHORS   Appel,B., Korzh,V., Glasgow,E., Thor,S., Edlund,T., Dawid,I.B. and
            Eisen,J.S.
  TITLE     Motoneuron fate specification and patterned LIM homeobox gene
            expression in embryonic Zebrafish
  JOURNAL   Development (1995) In press
REFERENCE   2  (bases 1 to 1369)
  AUTHORS   Glasgow,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-AUG-1995) Eric Glasgow, Laboratory of Molecular
            Genetics, NIH/NICHD, 9000 Rockville Pike, Bethesda, MD 20892, USA
FEATURES             Location/Qualifiers
     source          1. .1369
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
                     /dev_stage="embryo"
     gene            144. .1340
                     /gene="lim3"
     CDS             144. .1340
                     /gene="lim3"
```

```
                     /codon_start=1
                     /product="LIM homeobox protein"
                     /protein_id="AAA76714.1"
                     /db_xref="PID:g1000342"
                     /db_xref="GI:1000342"
                     /translation="MLLEHPGSSCQNAGNYTRYSSSQDIPVCAGCNQHIVDRFILKVL
                     DRHWHSKCLKCSDCQSQLADKCFSRGDSVYCKDDFFKRFGTKCAACQQGIPPTQVVRR
                     AQDFVYHLHCFACIVCKRQLATGDEYYLMEDSRLVCKADYETAKQREADSTAKRPRTT
                     ITAKQLETLKNAYNNSPKPARHVREQLSTETGLDMRVVQVWFQNRRAKEKRLKKDAGR
                     QRWGQYFRNMKRSRGTSKSDKDSTQEDGMDSDAEVSFTDEPPMSDLGHSNGIYSSLSE
                     SSPALSRQGGNHPAFPLEHGAIIPSQEPYHDIQASSPYSLPQSPGPLQPLPRHQPLIS
                     SLVYPESGLPMAGQSGGQDMTPGVRMMAAGNGPSSDLSTGSSGGYPDFPASPASWLDE
                     VDHAQF"
BASE COUNT      333 a    390 c    388 g    258 t
ORIGIN


  Query Match             5.7%;  Score 32.8;  DB 4;  Length 1369;
  Best Local Similarity   59.8%;  Pred. No. 6.7;
  Matches   55;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy    172 TTCTTCCCACTCCGCCCAGACAATGAATCATCGGTAAATGCTTATCTGATGAACTACGCC 231
          | |||| |||||||| ||  | |  |  | || ||   |  || | || || | |   |
Db     49 TGCTTCACACTCCGCTCACTCCAGCACGCGTCTGTCTGTTTTTCTTTGCTGCAGAAACTC 108

Qy    232 CCGGCGCGCCATTTTGGTTACTATTTTTCACC 263
          | | | ||| | ||| ||| | ||| || ||
Db    109 CGGTCTCGCTAGTTTAGTTTATTTTTATCGCC 140


RESULT  15
AC000134
LOCUS       AC000134   203300 bp    DNA             PRI       06-MAY-1999
DEFINITION  Homo sapiens Chromosome 11q13 BAC Clone 137c7, complete sequence.
ACCESSION   AC000134
NID         g4755212
VERSION     AC000134.14  GI:4755212
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 203300)
  AUTHORS   Crabtree,J. and Roe,B.A.
  TITLE     sequence of a 11q13 bac mapping to PYGM
  JOURNAL   Unpublished (1997)
REFERENCE   2  (bases 1 to 203300)
  AUTHORS   Chandrasekharappa,S.C., Guru,S.C., Manickam,P., Olufemi,S.-E. and
            Collins,F.S.
  TITLE     Sequence of a 11q13 bac mapping to PYGM
  JOURNAL   Unpublished (1997)
REFERENCE   3  (bases 1 to 203300)
  AUTHORS   Emmert-Buck,M.R., Debelenko,L.V., Zhuang,Z., Lubensky,I.A. and
            Liotta,L.A.
  TITLE     Sequence of a 11q13 bac mapping to PYGM
  JOURNAL   Unpublished (1997)
REFERENCE   4  (bases 1 to 203300)
  AUTHORS   Weisemann,J. and Boguski,M.S.
  TITLE     Sequence of a 11q13 bac mapping to PYGM
  JOURNAL   Unpublished (1997)
REFERENCE   5  (bases 1 to 203300)
  AUTHORS   Agarwal,S.K., Kester,M., Kim,Y.S., Heppner,C., Dong,Q.,
            Spiegel,A.M., Burns,L.A. and Marx,S.J.
  TITLE     Sequence of a 11q13 bac mapping to PYGM
  JOURNAL   Unpublished (1997)
REFERENCE   6  (bases 1 to 203300)
  AUTHORS   Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-1997) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   7  (bases 1 to 203300)
  AUTHORS   Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-MAR-1999) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   8  (bases 1 to 203300)
  AUTHORS   Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-APR-1999) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   9  (bases 1 to 203300)
  AUTHORS   Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-APR-1999) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   10 (bases 1 to 203300)
  AUTHORS   Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAY-1999) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   11 (bases 1 to 203300)
  AUTHORS   Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAY-1999) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   12 (bases 1 to 203300)
  AUTHORS   Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAY-1999) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   13 (bases 1 to 203300)
  AUTHORS   Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAY-1999) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On May 6, 1999 this sequence version replaced gi:4753208.
FEATURES             Location/Qualifiers
     source          1. .203300
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="137c7"
BASE COUNT    49152 a   57085 c   53982 g   43081 t
ORIGIN


  Query Match             5.7%;  Score 32.8;  DB 42;  Length 203300;
  Best Local Similarity   51.6%;  Pred. No. 10;
  Matches   99;  Conservative    0;  Mismatches   92;  Indels    1;  Gaps    1;

Qy    381 ATTGGTCTATCGCTTCTCCCTGACTGGCTGAGCTCTCTTTTTGGCTGTTGGTATGCACCT 440
          |||  |||||| | | |    ||| | | | | |||||   | ||| | ||    |
Db 165943 ATTCCTCTATCCATCCCTCAACCCTGCTTTATATTTTTTTTTCTTTTTGTTTTGTTTTT 166002

Qy    441 CGCCAAAGGTGTAGCTCCCTCTGAAATAGGTGGTAATTGTTTTGC-CTGCATCIGATCTT 499
          |   || || ||  | |||| | | | |||  | | |     | || |  | |  | ||
Db 166003 CTGAAACGGAGTCTCACCCTGTCACACAGGCTGGAGTGCAGTGGCTCGACCTTGGCTCAC 166062

Qy    500 CCGACGTTAACACCACCAGGCACGAGCATTCTTTTTCAAGAAGCATTTTCATATGCGCTT 559
          |  |   |    | |||||  | |||| |  |  |       || | ||||  ||| |||
Db 166063 CGCAACCTCTGCCTCCCAGGTTCAAGCAATTCTCCTGCCTCGGCCTCTTCAGGTGCACTT 166122

Qy    560 CCAGCGCATCCC 571
          |  | |||  ||
Db 166123 CAGGTGCACGCC 166134


Search completed: October  5, 1999, 18:04:03
```

Job time: 11686 sec

OM of: US-09-068-804-8 to: EST:* out_format : pfs

Date: Oct 5, 1999 7:34 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09068804/runat_05101999_145134_10306/app_query.fasta.1
-DB=EST -QFMT=fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=pam150 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=escore -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfs -NORM=stat -USER=US09068804 -NCPU=6 -ICPU=3 -WAIT
-THREADS=1

Search information block:
Query: US-09-068-804-8
Query length: 64
Database: EST:*
Database sequences: 2546578
Database length: 986266752
Search time (sec): 940.790000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ..
gb_est20:AA852614	-	63.00	168.78	0.6918	308	! AA852614 NHTBCae14g08f1 Normal
gb_est18:AA675690	+	63.00	162.45	1.56	550	! AA675690 vp96a10.r1 Stratagene
gb_est15:AA472587	+	61.50	159.33	2.33	495	! AA472587 vh04b10.r1 Soares mous
gb_est9:AA105333	+	61.50	158.50	2.59	534	! AA105333 mp37c09.r1 Barstead MF
gb_est29:AI549972	+	59.00	152.27	5.75	492	! AI549972 mm95e03.y1 Stratagene
gb_est9:AA073989	+	57.50	151.58	6.28	354	! AA073989 mm95e03.r1 Stratagene
gb_est26:AI373032	-	58.50	151.55	6.31	461	! AI373032 qz02h08.x1 NCI_CGAP_Br
gb_est23:AI118721	-	58.50	151.13	6.65	479	! AI118721 uc12g01.r1 Soares mous
gb_est24:AI222444	-	57.50	149.06	8.68	446	! AI222444 qh04h09.x1 Soares_NFL_
gb_est30:AI633006	-	57.00	148.09	9.84	428	! AI633006 tz33c11.x1 NCI_CGAP_Ut
gb_est23:AI099303	+	57.00	147.76	10.26	441	! AI099303 ue33d04.y1 Sugano mous
gb_est21:C91744	-	56.50	146.87	11.50	420	! C91744 C91744 Rice panicle shor
gb_est31:AI683338	-	57.00	146.74	11.68	484	! AI683338 tw50h07.x1 NCI_CGAP_Ut
gb_est23:AI098937	+	57.00	146.52	12.02	494	! AI098937 ue34a04.y1 Sugano mous
gb_est25:AI313959	+	57.00	146.37	12.26	501	! AI313959 uj49c06.y1 Sugano mous
gb_est31:AI688425	+	55.00	144.90	14.80	340	! AI688425 wc94h12.x1 NCI_CGAP_Co
gb_est2:T62409	-	54.00	143.90	16.83	287	! T62409 tgEST0173 TgRH Tachyzoit
gb_est9:AA106293	+	56.00	142.78	19.43	536	! AA106293 mm18f12.r1 Stratagene
gb_est6:W11298	-	53.50	142.59	19.91	284	! W11298 ma79b03.r1 Soares mouse
gb_est10:AA152485	+	54.50	141.95	21.62	391	! AA152485 zo30g09.r1 Stratagene
gb_est20:AA895166	+	53.00	141.83	21.93	267	! AA895166 vx47d06.r1 Stratagene
gb_est2:F00049	+	52.50	141.76	22.15	236	! F00049 HSB09D041 STRATAGENE Hum
gb_est16:C43314	-	54.00	141.42	23.12	360	! C43314 C43314 Yuji Kohara unpub
gb_est18:AA674266	+	55.00	140.95	24.55	488	! AA674266 vp98d02.r1 Stratagene
gb_est23:AI097749	+	57.00	140.67	25.45	844	! AI097749 ue39d06.y1 Sugano mous
gb_est19:AA759231	+	54.00	140.14	27.26	405	! AA759231 ah84e05.s1 Soares_test
gb_est32:AU068953	-	54.00	140.08	27.45	407	! AU068953 AU068953 Rice callus C
gb_est8:AA046596	-	53.00	139.86	28.26	320	! AA046596 zk62h05.r1 Soares_preg
gb_est31:AU036836	-	53.50	139.82	28.40	366	! AU036836 AU036836 Cryptomeria j
gb_est26:AI387889	+	55.00	139.73	28.74	546	! AI387889 GH18573.5prime GH Dros
gb_est28:AI529187	+	54.00	139.56	29.36	427	! AI529187 ui66d07.y1 Sugano mous
gb_est29:AI573781	+	57.00	139.35	30.17	953	! AI573781 uj65f10.y1 Sugano mous
gb_est17:AA620609	+	54.00	139.13	31.01	444	! AA620609 af84a10.s1 Soares_test
gb_est25:AI316482	+	57.00	139.13	31.01	972	! AI316482 uj60g12.y1 Sugano mous
gb_est8:AA023093	+	54.00	139.01	31.50	449	! AA023093 mh67c11.r1 Soares mous
gb_est19:AA753018	-	54.00	138.99	31.60	450	! AA753018 97AS1524 Rice Immature
gb_est23:AI155311	-	53.00	138.79	32.42	353	! AI155311 ud84e07.r1 Soares mous
gb_est16:AA562852	-	53.50	138.74	32.62	404	! AA562852 vl64e10.r1 Knowles Sol
gb_est18:AA668858	-	53.00	138.30	34.49	369	! AA668858 aa80e04.s1 NCI_CGAP_GC
gb_est18:AA711636	-	54.00	138.24	34.79	482	! AA711636 vu31h10.r1 Stratagene
gb_est23:AI082749	-	48.50	138.20	34.94	115	! AI082749 ow77g12.s1 Soares_feta
gb_est14:AA421091	+	53.00	138.16	35.15	374	! AA421091 zu05a02.r1 Soares_test
gb_est25:AI258633	+	55.00	138.11	35.35	633	! AI258633 LP01868.5prime LP D
gb_est14:AA421090	+	53.00	137.39	38.75	401	! AA421090 zu05a01.r1 Soares_t
gb_est22:AI064684	-	54.50	137.20	39.74	604	! AI064684 GH02917.5prime GH D

seq_name: gb_est20:AA852614

seq_documentation_block:
LOCUS AA852614 308 bp mRNA EST 12-MAY-1999
DEFINITION NHTBCae14g08f1 Normal Human Trabecular Bone Cells Homo sapiens cDNA clone NHTBCae14g08, mRNA sequence.
ACCESSION AA852614
NID g2941207
VERSION AA852614.1 GI:2941207
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 308)
AUTHORS Jia,L.B., Young,M.F., Touchman,J.W., Bouffard,G.G., Beckstrom-Sternberg,S.M., Green,E.D., Powell,J.I., Yang,L.M., Robey,P.G. and Francomano,C.A.
TITLE SGAP: The Skeletal Genome Anatomy Project
JOURNAL Unpublished (1997)
COMMENT On Jan 14, 1998 this sequence version replaced gi:1797183.

Contact: Libin Jia
Medical Genetics Branch
National Human Genome Research Institute
10/3D53, 9000 Rockville Pike, Bethesda, MD 20892-1267, USA
Tel: 301-496-1681
Fax: 301-435-3495
Email: libin@helix.nih.gov
Seq primer: M13 Forward.

FEATURES Location/Qualifiers
source 1. .308
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="10 p11.2-cen"
/clone="NHTBCae14g08"
/clone_lib="Normal Human Trabecular Bone Cells"
/sex="Female"
/tissue_type="Bone"
/cell_type="Trabecular Bone Cells"
/lab_host="SURE"
/note="Organ: Hip; Vector: pBluescript; Site_1: EcoRI"

BASE COUNT 57 a 78 c 109 g 63 t 1 others
ORIGIN

alignment_scores:
Quality: 63.00 Length: 45
Ratio: 1.800 Gaps: 1
Percent Similarity: 77.778 Percent Identity: 28.889

alignment_block:
US-09-068-804-8 x AA852614/rev ..

Align seg 1/1 to reverse of: AA852614 from: 1 to: 308

```
      9 ProProValIleMetProGlyMetGlnThrGluIleLysThrGlnAlaTh 25
        ||||||   |||   |||:::::::::|||:::::   |||::::::
    164 CCTCCCTGTATCACCCCCTCCCTACACACACACACACACACACACACA 115

     25 rAsnLeuAlaAlaAsnLeuSerAla.ValArgGluSerAlaThrAlaThr 41
         :::|||:::::   |||:::::: :::|||:::::::::
    114 CACACTGCCTCCCATCCTTCCCTCATGCCCGCCAGTGCACAGGGAAGGGC 65

     42 LeuSerGlyGluIleLysGlyProGlnLeuGlu 52
        |||:::::::::::::   |||:::::||||||
     64 TTGGCCAGCGCTGTTGAGGGGTCCCCTCTGGAA 32
```

seq_name: gb_est18:AA675690

```
seq_documentation_block:
LOCUS       AA675690      550 bp    mRNA            EST       28-NOV-1997
DEFINITION  vp96a10.r1 Stratagene mouse diaphragm (#937303) Mus musculus cDNA
            clone IMAGE:1092570 5' similar to gb:M12529 APOLIPOPROTEIN E
            PRECURSOR (HUMAN); gb:D00466 Mouse apolipoprotein E gene (MOUSE);,
            mRNA sequence.
ACCESSION   AA675690
NID         g2651349
VERSION     AA675690.1  GI:2651349
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 550)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1395380.

            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:598802
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 338.
FEATURES             Location/Qualifiers
     source          1. .550
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1092570"
                     /clone_lib="Stratagene mouse diaphragm (#937303)"
                     /tissue_type="diaphragm"
                     /dev_stage="adult"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Organ: diaphragm; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally from mRNA
                     prepared from diaphragm muscle.  Primer: Oligo dT. Average
                     insert size: 1.5 kb. Uni-ZAP XR Vector; ~5' adaptor
                     sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor sequence: 5'
                     CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT      147 a    138 c    182 g     83 t
ORIGIN


alignment_scores:
            Quality:   63.00            Length:     54
              Ratio:  1.658              Gaps:      1
 Percent Similarity:  70.370   Percent Identity: 29.630

alignment_block:
 US-09-068-804-8 x AA675690   ..

 Align seg 1/1  to: AA675690  from: 1  to: 550

      12 IleMetProGlyMetGlnThrGluIleLysThrGlnAlaThrAsnLeuAl 28
         ::|||   :::      ||||||:::|||:::      :::::::|||::
     239 CTGATGGAGGACACTATGACGGAAGTAAAGGCTTACAAAAAGGAGCTGGA 288

      28 aAlaAsnLeuSerAlaValArgGluSerAlaThrAlaThrLeuSerGlyG 45
         ::::::::|||:::::::|||   |||   :::||||||   |||:::   |
     289 GGAACAGCTGGGTCCAGTGGCGGAGGAGACAACGGCCAGGCTGGGCAAAG 338

      45 luIleLysGlyProGln.................LeuGluAspPhePro 55
         ||::::::::::::|||                  :::||||||:::
     339 AGGTGCAGGCTGCACAAGGCCGACTCCGAGCCGACATGGAGGATCTACGC 388

      56 AlaLeuIleLys 59
         :::|||::::::
     389 AACCTGACTCGG 400

seq_name: gb_est15:AA472587

seq_documentation_block:
LOCUS       AA472587      495 bp    mRNA            EST       18-JUN-1997
DEFINITION  vh04b10.r1 Soares mouse mammary gland NbMMG Mus musculus cDNA clone
            IMAGE:874459 5', mRNA sequence.
ACCESSION   AA472587
NID         g2200869
VERSION     AA472587.1  GI:2200869
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 495)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1407318.

            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:513939
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 414.
FEATURES             Location/Qualifiers
     source          1. .495
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:874459"
                     /clone_lib="Soares mouse mammary gland NbMMG"
                     /sex="male"
                     /tissue_type="mammary gland"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
                     /note="Organ: mammary gland; Vector: pT7T3D-Pac
                     (Pharmacia) with a modified polylinker; Site_1: Not I;
                     Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                     oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATGGTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
                     constructed and normalized by Bento Soares and M.Fatima
                     Bonaldo."
BASE COUNT      105 a    151 c    155 g     84 t
ORIGIN


alignment_scores:
            Quality:   61.50            Length:     64
```

```
            Ratio:  1.367              Gaps:      4
 Percent Similarity:  70.312   Percent Identity: 32.812

alignment_block:
 US-09-068-804-8 x AA472587   ..

 Align seg 1/1  to: AA472587  from: 1  to: 495

       1 MetValThrSerValArgThrGlnProProValIleMetProGlyMetGl 17
         :::::::::::::||||||:::|||||   :::      ||||||
     287 GTCATTAGCACGGTGCGGCCACAGCCCCACCTGGCAGAGCCAGGA..... 331

      17 nThrGluIleLysThrGlnAlaThrAsnLeuAlaAlaAsnLeuSerAlaV 34
                 |||:::|||:::      |||:::::   |||:::   :
     332 .........AAGGTGCAGACCCAGTTCCTCCCGGAGCTCCTGGGCCTTC 371

      34 alArgGluSerAlaThrAlaThrLeuSerGlyGluIleLysGly...Pro 49
         ::|||||:::::   |||   :::      :::|||:::::   |||
     372 TCCGTGAGCCTGGGGGAGCTCAAGTCCATCCGCAGATCCAAGCCAGGCCT 421

      50 GlnLeuGluAspPheProAla.LeuIleLysGlnAlaSer 62
         ||||||   ::::::|||::: |||:::::::||||||:::
     422 CAGCTG...GGCCTACCTGGTCCTGGTGACCCAGGCTGGA 458

seq_name: gb_est9:AA105333

seq_documentation_block:
LOCUS       AA105333      534 bp    mRNA            EST       13-FEB-1997
DEFINITION  mp37c09.r1 Barstead MPLRB1 Mus musculus cDNA clone IMAGE:571408 5',
            mRNA sequence.
ACCESSION   AA105333
NID         g1655165
VERSION     AA105333.1  GI:1655165
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 534)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1297484.

            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:346056
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 503.
FEATURES             Location/Qualifiers
     source          1. .534
                     /organism="Mus musculus"
                     /strain="BALB/c"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:571408"
                     /clone_lib="Barstead MPLRB1"
                     /sex="mixed"
                     /tissue_type="Kidney"
                     /dev_stage="6 weeks"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     [CATGGATTCGGTACC], digested with Not I and cloned into the
                     Not I and Eco RI sites of the modified pI7T3 vector.
                     Library constructed by Bob Barstead."
BASE COUNT      110 a    169 c    159 g     95 t      1 others
ORIGIN


alignment_scores:
            Quality:   61.50             Length:     64
              Ratio:  1.367                Gaps:      4
 Percent Similarity:  70.312   Percent Identity: 32.812

alignment_block:
 US-09-068-804-8 x AA105333   ..

 Align seg 1/1  to: AA105333  from: 1  to: 534

       1 MetValThrSerValArgThrGlnProProValIleMetProGlyMetGl 17
         :::::::::::::||||||:::||||||   :::      ||||||
     265 GTCATTAGCACGGTGCGGCCACAGCCCCACCTGGCAGAGCCAGGA..... 309

      17 nThrGluIleLysThrGlnAlaThrAsnLeuAlaAlaAsnLeuSerAlaV 34
                 |||:::|||:::      |||:::::   |||:::   :
     310 .........AAGGTGCAGACCCAGTTCCTCCCGGAGCTCCTGGGCCTTC 349

      34 alArgGluSerAlaThrAlaThrLeuSerGlyGluIleLysGly...Pro 49
         ::|||||:::::   |||   :::      :::|||:::::   |||
     350 TCCGTGAGCCTGGGGGAGCTCAAGTCCATCCGCAGATCCAAGCCAGGCCT 399

      50 GlnLeuGluAspPheProAla.LeuIleLysGlnAlaSer 62
         ||||||   ::::::|||::: |||:::::::||||||:::
     400 CAGCTG...GGCCTACCTGGTCCTGGTGACCCAGGCTGGA 436

seq_name: gb_est29:AI549972

seq_documentation_block:
LOCUS       AI549972      492 bp    mRNA            EST       23-MAR-1999
DEFINITION  mm95e03.y1 Stratagene mouse heart (#937316) Mus musculus cDNA clone
            IMAGE:536188 5' similar to TR:P70617 P70617 NINJURIN1. ;, mRNA
            sequence.
ACCESSION   AI549972
NID         g4482335
VERSION     AI549972.1  GI:4482335
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 492)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
            Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
            Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3189647.

            Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            This read is a RESEQUENCE of a previously sequenced mouse clone
            This read has been verified (found to hit its original self in the
            correct orientation)
            Seq primer: -40RP from Gibco
```

```
          High quality sequence stop: 416.
FEATURES             Location/Qualifiers
     source          1. .492
                     /organism="Mus musculus"
                     /strain="NIH/Swiss"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:536188"
                     /clone_lib="Stratagene mouse heart (#937316)"
                     /sex="pooled"
                     /tissue_type="heart"
                     /dev_stage="13 day embryos"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Organ: heart; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. 93 pooled NIH/Swiss 13 day embryo hearts.
                     Average insert size: 1.0 kb; Uni-ZAP XR Vector; ~5'
                     adaptor sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor
                     sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT       91 a    152 c    154 g     95 t
ORIGIN


alignment_scores:
            Quality:   59.00             Length:     48
              Ratio:  1.686               Gaps:      2
 Percent Similarity:  72.917   Percent Identity: 31.250

alignment_block:
 US-09-068-804-8 x AI549972   ..

 Align seg 1/1  to: AI549972  from: 1  to: 492

      14 ProGlyMetGlnThr...GluIleLysThrGlnAlaThrAsnLeuAlaAl 29
         ||||||   :::|||   |||      |||::::::   :::|||:::::
      24 CCGGGCGGCCGCACCATGGAGTCGGGCACTGAGGAGTATGAGCTCAACGG 73

      29 aAsnLeuSerAlaValArgGluSerAlaThrAlaThrLeuSerGlyGluI 46
         ::::|||   :::        ::::::::::::::|||        :
      74 CGACCTGCGCCCGGGCTCCCCGGGTTCCCCCGACGCCTTG.........G 114

      46 leLysGlyProGlnLeuGluAspPheProAlaLeuIleLysGln 60
         ::::::|||:::   ||||||:::||||||::::::||||:::|||
     115 TCACTGGCACAAGCCTGGAAGAGTTTCCTAGCATGATCAGGCAG 158

seq_name: gb_est9:AA073989

seq_documentation_block:
LOCUS        AA073989      354 bp    mRNA            EST       15-FEB-1997
DEFINITION   mm95e03.r1 Stratagene mouse heart (#937316) Mus musculus cDNA clone
             IMAGE:536188 5', mRNA sequence.
ACCESSION    AA073989
NID          g1595029
VERSION      AA073989.1  GI:1595029
KEYWORDS     EST.
SOURCE       house mouse.
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
             Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 354)
  AUTHORS    Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
             Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
             Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
             Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
             Waterston,R.
  TITLE      The WashU-HHMI Mouse EST Project
  JOURNAL    Unpublished (1996)
COMMENT      On Sep 12, 1996 this sequence version replaced gi:1395280.

             Contact: Marra M/Mouse EST Project
             WashU-HHMI Mouse EST Project
             Washington University School of MedicineP
             4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
             Tel: 314 286 1800
             Fax: 314 286 1810
             Email: mouseest@watson.wustl.edu
             This clone is available royalty-free through LLNL ; contact the
             IMAGE Consortium (info@image.llnl.gov) for further information.
             MGI:323124
             Seq primer: -28m13 rev1 ET from Amersham
             High quality sequence stop: 350.
FEATURES             Location/Qualifiers
     source          1. .354
                     /organism="Mus musculus"
                     /strain="NIH/Swiss"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:536188"
                     /clone_lib="Stratagene mouse heart (#937316)"
                     /sex="pooled"
                     /tissue_type="heart"
                     /dev_stage="13 day embryos"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Organ: heart; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. 93 pooled NIH/Swiss 13 day embryo hearts.
                     Average insert size: 1.0 kb; Uni-ZAP XR Vector; ~5'
                     adaptor sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor
                     sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT       69 a    110 c    111 g     62 t      2 others
ORIGIN


alignment_scores:
            Quality:   57.50             Length:     39
              Ratio:  1.917               Gaps:      1
 Percent Similarity:  76.923   Percent Identity: 28.205

alignment_block:
 US-09-068-804-8 x AA073989   ..

 Align seg 1/1  to: AA073989  from: 1  to: 354

      22 ThrGlnAlaThrAsnLeuAlaAlaAsnLeuSerAlaValArgGluSerAl 38
         |||::::::   :::|||::::::::::|||   :::         :::::
      51 ACTGAGGAGTATGAGCTCAACGGCGACCTGCGCCCGGGCTCCCCGGGTTC 100

      38 aThrAlaThrLeuSerGlyGluIleLysGlyProGlnLeuGluAspPheP 55
         ::::::::::|||         ::::::|||:::   ||||||:::||||
     101 CCCCGACGCCTTG.........GTCACTGGCACAAGCCTGGAAGAGTTTC 141

      55 roAlaLeuIleLysGln 60
         ||:::::::|||:::|||
     142 CTAGCATGATCAGGCAG 158

seq_name: gb_est26:AI373032

seq_documentation_block:
LOCUS        AI373032      461 bp    mRNA            EST       16-FEB-1999
DEFINITION   qz02h08.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2020383 3',
             mRNA sequence.
ACCESSION    AI373032
NID          g4152898
VERSION      AI373032.1  GI:4152898
KEYWORDS     EST.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
             Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 461)
  AUTHORS    NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE      National Cancer Institute / National Institute of Neurological
             Disorders and Stroke, Brain Tumor Genome Anatomy Project
             (CGAP/BTGAP), Tumor Gene Index
  JOURNAL    Unpublished (1998)
COMMENT      On May 1, 1998 this sequence version replaced gi:3106364.

             Contact: Robert Strausberg, Ph.D.
```

```
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
            Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 1607   Std Error: 0.00
            Seq primer: -40UP from Gibco.
FEATURES              Location/Qualifiers
     source           1. .461
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:2020383"
                      /clone_lib="NCI_CGAP_Brn25"
                      /tissue_type="anaplastic oligodendroglioma"
                      /lab_host="DH10B"
                      /note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
                      modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                      strand cDNA was primed with a Not I - oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCATAGGTTTTTTTTTTTTTTTTTTTTTT
                      T 3']; double-stranded cDNA was ligated to Eco RI
                      adaptors (Pharmacia), digested with Not I and cloned into
                      the Not I and Eco RI sites of the modified pT7T3 vector.
                      Library is normalized, and was constructed by Bento
                      Soares and M.Fatima Bonaldo."
BASE COUNT      127 a    108 c    124 g    102 t
ORIGIN


alignment_scores:
            Quality:   58.50            Length:     60
              Ratio:  1.330              Gaps:      2
 Percent Similarity:  73.333   Percent Identity: 26.667

alignment_block:
 US-09-068-804-8 x AI373032/rev   ..

 Align seg 1/1  to reverse of: AI373032  from: 1  to: 461

       5 ValArgThrGlnProProValIleMetProGlyMetGlnThrGluIleLy 21
         ::::::::::|||   |||:::      |||         ::::::::
     307 CTCAAAGTCCAGCACCCTGCGGGGCAGCCCCTGAGGGCCATCCAGCTTCC 258

      21 sThrGlnAlaThrAsnLeuAlaAlaAsnLeuSerAlaValArgGluSerA 38
          :::::::::|||   :::::::::   ||||||      |||::::::
     257 ATCTCCCAACACTAGGTTCACTTCGTGTCTCTCA......CGACAGGGAA 214

      38 laThrAlaThrLeuSerGlyGluIleLysGly...ProGlnLeuGluAsp 53
           :::   ::::::   |||::::::||||||   :::::::|||   :::
     213 GAAATTTAGCATTCCTTGGAGACCTGAAAGGATGCAGTGAGCTTAAGAAT 164

      54 PheProAlaLeuIleLysGlnAlaSerLeu 63
         |||::::::||||||:::|||::::::|||
     163 TTTCAAGAGCTTATCAATCAGTCAGCCCTT 134

seq_name: gb_est23:AI118721

seq_documentation_block:
LOCUS       AI118721      479 bp    mRNA            EST       02-SEP-1998
DEFINITION  uc12g01.r1 Soares mouse mammary gland NbMMG Mus musculus cDNA clone
            IMAGE:1397808 5' similar to TR:O35494 O35494 FRIZZLED-9 PROTEIN ;,
            mRNA sequence.
ACCESSION   AI118721
NID         g3519045
VERSION     AI118721.1  GI:3519045
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 479)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2152103.

            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:909524
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 453.
FEATURES              Location/Qualifiers
     source           1. .479
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /map="10"
                      /clone="IMAGE:1397808"
                      /clone_lib="Soares mouse mammary gland NbMMG"
                      /sex="male"
                      /tissue_type="mammary gland"
                      /dev_stage="4 weeks"
                      /lab_host="DH10B"
                      /note="Organ: mammary gland; Vector: pT7T3D-Pac
                      (Pharmacia) with a modified polylinker; Site_1: Not I;
                      Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                      oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATGGTTTTTTTTTTTTTTTTTTTT
                      T 3']; double-stranded cDNA was ligated to Eco RI
                      adaptors (Pharmacia), digested with Not I and cloned into
                      the Not I and Eco RI sites of the modified pT7T3 vector.
                      RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
                      constructed and normalized by Bento Soares and M.Fatima
                      Bonaldo."
BASE COUNT      105 a    136 c    132 g    105 t      1 others
ORIGIN


alignment_scores:
            Quality:   58.50            Length:     49
              Ratio:  1.828              Gaps:      2
 Percent Similarity:  65.306   Percent Identity: 34.694

alignment_block:
 US-09-068-804-8 x AI118721/rev   ..

 Align seg 1/1  to reverse of: AI118721  from: 1  to: 479

       2 ValThrSerValArgThrGlnProProValIleMetPro..... 14
         |||      ::::::||||||::::|||::::::||||||
     331 GTCTTGGAACTCCATACCCAGACTCCACTGGTGATGCCCACCACCAAGGA 282

      15 .............GlyMetGlnThrGluIleLysThrGlnAlaThrAsnL 27
                      :::|||:::|||:::       |||   :::   |
     281 CATGAAGATTTTGAGCATGAAGACAGACCACAGTGGGCAACGTAGCCCCC 232

      27 euAlaAla...AsnLeuSerAlaValArgGluSerAlaThrAlaThr 41
         ||||||||   :::|||||||||   |||:::::::::::::|||
     231 TGGCAGCGAGCAGTCTCTCGGCCTCCAGGCACGGTGGCAGCAGTACA 185
```

```
seq_name: gb_est24:AI222444

seq_documentation_block:
LOCUS       AI222444      446 bp    mRNA            EST       30-NOV-1998
DEFINITION  qh04h09.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
            IMAGE:1843745 3', mRNA sequence.
ACCESSION   AI222444
NID         g3804647
VERSION     AI222444.1  GI:3804647
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 446)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Aug 21, 1998 this sequence version replaced.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 716   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 403.
FEATURES             Location/Qualifiers
     source          1. .446
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1843745"
                     /clone_lib="Soares_NFL_T_GBC_S1"
                     /lab_host="DH10B"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from three normalized
                     libraries (fetal lung NbHL19W, testis NHT, and B-cell
                     NCI_CGAP_GCB1) were mixed, and ss circles were made in
                     vitro. Following HAP purification, this DNA was used as
                     tracer in a subtractive hybridization reaction. The driver
                     was PCR-amplified cDNAs from pools of 5,000 clones made
                     from the same 3 libraries. The pools consisted of
                     I.M.A.G.E. clones 297480-302087, 682632-687239,
                     726408-728711, and 729096-731399. Subtraction by Bento
                     Soares and M. Fatima Bonaldo.  "
BASE COUNT       99 a     92 c     85 g    170 t
ORIGIN


alignment_scores:
            Quality:   57.50            Length:     76
              Ratio:  1.223              Gaps:      4
 Percent Similarity:  61.842   Percent Identity: 23.684

alignment_block:
 US-09-068-804-8 x AI222444/rev   ..

 Align seg 1/1  to reverse of: AI222444  from: 1  to: 446

      1 MetValThrSerValArgThrGlnPro.....................Pr 10
        |||   |||:::|||::::::::::|||                    ||
    273 ATGGAAACTCCAGTAAAAAAAGAACCAAGTGATGACTTAAGAGAACTTCC 224

     10 oVal..............IleMetProGlyMet................ 16
        ||||              |||:::|||::::::
    223 TGTACTGGACTGTGGATGGATAAAGCCAGACATCTGTGCTTCCAACTCAG 174

     17 ..GlnThrGluIleLysThrGlnAlaThrAsnLeuAlaAlaAsnLeuSer 32
         ::::::|||||||||   :::::::::::::::::::::::::|||:::
    173 CAGAAAGTGAAATAAAACGTGATCCCAAAGATGTAAACACTTCCTTAGGA 124

     33 AlaValArgGluSerAlaThr.............AlaThrLeuSerGl 44
         :::|||   :::   ::::::                 :::   :::|||
    123 GAAGTTGCTAATGAGACCTCTGAAAATGAAACACTGGGAGACTTCAGTGA 74

     44 yGluIleLysGlyProGlnLeuGluAsp 53
         :::||||||||||:::::::|||::::::
     73 ACAAATAAAAGGTTCAGACTTGGATGAA 46

seq_name: gb_est30:AI633006

seq_documentation_block:
LOCUS       AI633006      428 bp    mRNA            EST       26-APR-1999
DEFINITION  tz33c11.x1 NCI_CGAP_Ut2 Homo sapiens cDNA clone IMAGE:2290388 3',
            mRNA sequence.
ACCESSION   AI633006
NID         g4684336
VERSION     AI633006.1  GI:4684336
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 428)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On May 18, 1998 this sequence version replaced gi:3137095.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco
            High quality sequence stop: 415.
FEATURES             Location/Qualifiers
     source          1. .428
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2290388"
                     /clone_lib="NCI_CGAP_Ut2"
                     /tissue_type="moderately-differentiated endometrial
                     adenocarcinoma, 3 pooled tumors"
                     /lab_host="DH10B"
                     /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.85 kb. Life Technologies catalog #:
                     11539-012"
BASE COUNT      121 a     99 c    115 g     93 t
ORIGIN


alignment_scores:
            Quality:   57.00            Length:     57
              Ratio:  1.390              Gaps:      3
 Percent Similarity:  71.930   Percent Identity: 29.825

alignment_block:
 US-09-068-804-8 x AI633006/rev   ..

 Align seg 1/1  to reverse of: AI633006  from: 1  to: 428

      8 GlnProProValIleMetProGlyMetGlnThrGluIleLysThrGlnAl 24
```

```
      |||||||||      |||              ::::::   ::::::::
  281 CAGCCCCCAAGGGCCATG.............CAGCTTCTGTCTCCCAA 247

   24 aThrAsnLeuAlaAlaAsnLeuSerAlaValArgGluSerAlaThrAlaT 41
      :|||:::::::::::   ||||||     |||:::   ::::::   :
  246 CACTAAGTTCACTTCGTGTCTCTCA......CGGCAGAGAGGAAACTTAG 203

   41 hrLeuSerGlyGluIleLysGly...ProGlnLeuGluAspPheProAla 56
      :::::   |||::::::|||||   :::::|||   :::|||::::::
  202 TATTCCTTGGAGACCTGAAGGGATGCAGTGAGCTTAAGAATTTTCAAGAG 153

   57 LeuIleLysGlnAlaSerLeu 63
      ||||||::::|||::::::|||
  152 CTTATCAATCAGTCAGCCCTT 132

seq_name: gb_est23:AI099303

seq_documentation_block:
LOCUS       AI099303      441 bp    mRNA            EST       20-AUG-1998
DEFINITION  ue33d04.y1 Sugano mouse liver mlia Mus musculus cDNA clone
            IMAGE:1482151 5' similar to gb:D00466 Mouse apolipoprotein E gene
            (MOUSE);, mRNA sequence.
ACCESSION   AI099303
NID         g3448828
VERSION     AI099303.1  GI:3448828
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 441)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Sep 4, 1996 this sequence version replaced gi:1505360.

            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:930507
            Trace considered overall poor quality
            Seq primer: custom primer used
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .441
                     /organism="Mus musculus"
                     /strain="C57BL"
                     /db_xref="taxon:10090"
                     /map="Chromosome X; Xp11.2"
                     /clone="IMAGE:1482151"
                     /clone_lib="Sugano mouse liver mlia"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: liver; Vector: pME18S-FL3; Site_1: DraIII
                     (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
                     was primed with an oligo(dT) primer
                     [ATGTGGCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was
                     ligated to a  DraIII adaptor [TGTTGGCCTACTGG], digested
                     and cloned into distinct DraIII sites of the pME18S-FL3
                     vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should
                     be used to isolate the cDNA insert.  Size selection was
                     performed to exclude fragments <1.5kb.  Library
                     constructed by Dr. Sumio Sugano (University of Tokyo
                     Institute of Medical Science).  Custom primers  for
                     sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
                     primer CGACCTGCAGCTCGAGCACA."
BASE COUNT      115 a    115 c    147 g     64 t
ORIGIN


alignment_scores:
            Quality:   57.00             Length:     44
              Ratio:  1.727               Gaps:      0
 Percent Similarity:  75.000   Percent Identity: 31.818

alignment_block:
 US-09-068-804-8 x AI099303   ..

 Align seg 1/1  to: AI099303  from: 1  to: 441

    7 ThrGlnProProValIleMetProGlyMetGlnThrGluIleLysThrGl 23
      ||||||   ::::::::::|||   :::    ||||||:::|||:::
  221 ACACAACTGACGGCACTGATGGAGGACACTATGACGGAAGTAAAGGCTTA 270

   23 nAlaThrAsnLeuAlaAlaAsnLeuSerAlaValArgGluSerAlaThrA 40
      ::::::|||::::::::::|||:::::|||   |||   :::   |
  271 CAAAAAGGAGCTGGAGGAACAGCTGGGTCCAGTGGCGGAGGAGACACGGG 320

   40 laThrLeuSerGlyGluIleLysGlyProGln 50
      ||   |||:::   |||:::::::::::::|||
  321 CCAGGCTGGGCAAAGAGGTGCAGGCGGCACAA 352

seq_name: gb_est21:C91744

seq_documentation_block:
LOCUS       C91744        420 bp    mRNA            EST       20-APR-1998
DEFINITION  C91744 Rice panicle shorter than 3cm Oryza sativa cDNA clone
            E31454_2A, mRNA sequence.
ACCESSION   C91744
NID         g3061112
VERSION     C91744.1  GI:3061112
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1  (bases 1 to 420)
  AUTHORS   Sasaki,T. and Yamamoto,K.
  TITLE     Rice cDNA from panicle
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 9, 1998 this sequence version replaced gi:937103.

            Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program
            2-1-2 Kannondai,Tsukuba
            Ibaraki,
            Japan 305
            Tel: 0298-38-7441
            Fax: 0298-38-7468
            Email: tsasaki@abr.affrc.go.jp.
FEATURES             Location/Qualifiers
     source          1. .420
                     /organism="Oryza sativa"
                     /strain="Nipponbare"
                     /db_xref="taxon:4530"
                     /clone="E31454_2A"
                     /clone_lib="Rice panicle shorter than 3cm"
                     /dev_stage="shorter than 3cm"
                     /note="Organ: panicle"
BASE COUNT       72 a    122 c    130 g     90 t      6 others
ORIGIN
```

```
alignment_scores:
            Quality:   56.50              Length:     55
              Ratio:  1.766                Gaps:      1
 Percent Similarity:  58.182   Percent Identity: 25.455

alignment_block:
 US-09-068-804-8 x C91744/rev   ..

 Align seg 1/1  to reverse of: C91744  from: 1  to: 420

       8 GlnProProValIleMetProGlyMetGlnThrGluIleLysThrGlnAl 24
         ||||||      :::::::|||
     320 CAGCCACTAAGCCTACTCCCA.............................. 300

      24 aThrAsnLeuAlaAlaAsnLeuSerAlaValArgGluSerAlaThrAlaT 41
           :::|||||||:::   |||   ::::::   |||:::::::
     299 ....TCTTTAGCACCACTCCTTGAGAACATCCCAGAGGCTCCCGGTTGCG 254

      41 hrLeuSerGlyGluIleLysGlyProGlnLeuGluAspProAlaLeu 57
          |||::::::   |||:::::::|||:::   ::::::||||||
     253 GGCTCCCATCCTTGATCCGCGCGCCCCACCCACAGCAATTCCCCATCCCC 204

      58 IleLysGlnAlaSer 62
         ::::::||||||:::
     203 CTCCAACAAGCGACT 189

seq_name: gb_est31:AI683338

seq_documentation_block:
LOCUS       AI683338      484 bp    mRNA            EST       26-MAY-1999
DEFINITION  tw50h07.x1 NCI_CGAP_Ut1 Homo sapiens cDNA clone IMAGE:2263165 3',
            mRNA sequence.
ACCESSION   AI683338
NID         g4893520
VERSION     AI683338.1  GI:4893520
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 484)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On May 18, 1998 this sequence version replaced gi:3138049.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco
            High quality sequence stop: 411.
FEATURES             Location/Qualifiers
     source          1. .484
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="955F01; 16; 16p13"
                     /clone="IMAGE:2263165"
                     /clone_lib="NCI_CGAP_Ut1"
                     /tissue_type="well-differentiated endometrial
                     adenocarcinoma, 7 pooled tumors"
                     /lab_host="DH10B"
                     /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.75 kb. Life Technologies catalog #:
                     11538-014"
BASE COUNT      137 a    109 c    127 g    111 t
ORIGIN


alignment_scores:
            Quality:   57.00              Length:     57
              Ratio:  1.390                Gaps:      3
 Percent Similarity:  71.930   Percent Identity: 29.825

alignment_block:
 US-09-068-804-8 x AI683338/rev   ..

 Align seg 1/1  to reverse of: AI683338  from: 1  to: 484

       8 GlnProProValIleMetProGlyMetGlnThrGluIleLysThrGlnAl 24
         |||||||||      |||              ::::::   ::::::::
     281 CAGCCCCCAAGGGCCATG..............CAGCTTCTGTCTCCCAA 247

      24 aThrAsnLeuAlaAlaAsnLeuSerAlaValArgGluSerAlaThrAlaT 41
         :|||:::::::::::::   ||||||      |||:::   ::::::   :
     246 CACTAAGTTCACTTCGTGTCTCTCA......CGGCAGAGAGGAAACTTAG 203

      41 hrLeuSerGlyGluIleLysGly...ProGlnLeuGluAspProAla 56
         :::::   |||:::::::||||||   ::::::|||   :::|||:::::
     202 TATTCCTTGGAGACCTGAAGGGATGCAGTGAGCTTAAGAATTTTCAAGAG 153

      57 LeuIleLysGlnAlaSerLeu 63
         ||||||:::|||:::::|||
     152 CTTATCAATCAGTCAGCCCTT 132

seq_name: gb_est23:AI098937

seq_documentation_block:
LOCUS       AI098937      494 bp    mRNA            EST       20-AUG-1998
DEFINITION  ue34a04.y1 Sugano mouse liver mlia Mus musculus cDNA clone
            IMAGE:1482222 5' similar to gb:D00466 Mouse apolipoprotein E gene
            (MOUSE);, mRNA sequence.
ACCESSION   AI098937
NID         g3448462
VERSION     AI098937.1  GI:3448462
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 494)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Nov 29, 1993 this sequence version replaced gi:635983.

            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:930578
            Seq primer: custom primer used
            High quality sequence stop: 457.
FEATURES             Location/Qualifiers
     source          1. .494
                     /organism="Mus musculus"
                     /strain="C57BL"
```

```
                     /db_xref="taxon:10090"
                     /map="3"
                     /clone="IMAGE:1482222"
                     /clone_lib="Sugano mouse liver mlia"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: liver; Vector: pME18S-FL3; Site_1: DraIII
                     (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
                     was primed with an oligo(dT) primer
                     [ATGTGGCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was
                     ligated to a  DraIII adaptor [TGTTGGCCTACTGG], digested
                     and cloned into distinct DraIII sites of the pME18S-FL3
                     vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should
                     be used to isolate the cDNA insert.  Size selection was
                     performed to exclude fragments <1.5kb.  Library
                     constructed by Dr. Sumio Sugano (University of Tokyo
                     Institute of Medical Science).  Custom primers  for
                     sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
                     primer CGACCTGCAGCTCGAGCACA."
BASE COUNT      128 a    126 c    167 g     73 t
ORIGIN


alignment_scores:
            Quality:   57.00              Length:     39
              Ratio:  1.900                Gaps:      0
 Percent Similarity:  76.923   Percent Identity: 33.333

alignment_block:
 US-09-068-804-8 x AI098937   ..

 Align seg 1/1  to: AI098937  from: 1  to: 494

      12 IleMetProGlyMetGlnThrGluIleLysThrGlnAlaThrAsnLeuAl 28
         :::|||   :::      ||||||:::|||:::      ::::::|||::
     283 CTGATGGAGGACACTATGACGGAAGTAAAGGCTTACAAAAAGGAGCTGGA 332

      28 aAlaAsnLeuSerAlaValArgGluSerAlaThrAlaThrLeuSerGlyG 45
         :::::::|||::::::|||   |||   :::   ||||||||||:::   |
     333 GGAACAGCTGGGTCCAGTGGCGGAGGAGACACGGGCCACGCTGGGCAAAG 382

      45 luIleLysGlyProGln 50
         ||:::::::::::|||
     383 AGGTGCAGGCGGCACAG 399

seq_name: gb_est25:AI313959

seq_documentation_block:
LOCUS       AI313959      501 bp    mRNA            EST       17-DEC-1998
DEFINITION  uj49c06.y1 Sugano mouse liver mlia Mus musculus cDNA clone
            IMAGE:1923274 5' similar to gb:D00466 Mouse apolipoprotein E gene
            (MOUSE);, mRNA sequence.
ACCESSION   AI313959
NID         g4029086
VERSION     AI313959.1  GI:4029086
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 501)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Jan 14, 1998 this sequence version replaced gi:1798233.

            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:979566
            Seq primer: custom primer used
            High quality sequence stop: 316.
FEATURES             Location/Qualifiers
     source          1. .501
                     /organism="Mus musculus"
                     /strain="C57BL"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1923274"
                     /clone_lib="Sugano mouse liver mlia"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: liver; Vector: pME18S-FL3; Site_1: DraIII
                     (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
                     was primed with an oligo(dT) primer
                     [ATGTGGCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was
                     ligated to a  DraIII adaptor [TGTTGGCCTACTGG], digested
                     and cloned into distinct DraIII sites of the pME18S-FL3
                     vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should
                     be used to isolate the cDNA insert.  Size selection was
                     performed to exclude fragments <1.5kb.  Library
                     constructed by Dr. Sumio Sugano (University of Tokyo
                     Institute of Medical Science).  Custom primers  for
                     sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
                     primer CGACCTGCAGCTCGAGCACA."
BASE COUNT      131 a    130 c    163 g     77 t
ORIGIN


alignment_scores:
            Quality:   57.00              Length:     39
              Ratio:  1.900                Gaps:      0
 Percent Similarity:  76.923   Percent Identity: 33.333

alignment_block:
 US-09-068-804-8 x AI313959   ..

 Align seg 1/1  to: AI313959  from: 1  to: 501

      12 IleMetProGlyMetGlnThrGluIleLysThrGlnAlaThrAsnLeuAl 28
         :::|||   :::      ||||||:::|||:::      ::::::|||::
     270 CTGATGGAGGACACTATGACGGAAGTAAAGGCTTACAAAAAGGAGCTGGA 319

      28 aAlaAsnLeuSerAlaValArgGluSerAlaThrAlaThrLeuSerGlyG 45
         :::::::|||::::::|||   |||   :::   ||||||||||:::   |
     320 GGAACAGCTGGGTCCAGTGGCGGAGGAGACACAGGCCACTCTGTGCAAAG 369

      45 luIleLysGlyProGln 50
         ||:::::::::::|||
     370 AGGTGCAGGCGGCACAA 386
```

OM of: US-09-068-804-8 to: Issued_Patents_NA:* out_format : pfs

Date: Oct 5, 1999 7:12 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09068804/runat_05101999_145135_10326/app_query.fasta.1
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=pam150
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=escore
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09068804
-NCPU=6 -ICPU=3 -WAIT -THREADS=1

Search information block:
Query: US-09-068-804-8
Query length: 64
Database: Issued_Patents_NA:*
Database sequences: 192659
Database length: 52021692
Search time (sec): 101.650000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ..
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-347-221-1	+	59.00	138.12	2.65	1400	!
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-040-548-32	+	52.00	119.99	27.10	1200	!
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-466-344-32	+	52.00	119.99	27.10	1200	!
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-233-608-6	-	49.00	119.97	27.18	540	! S
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-887-480-6	-	49.00	119.97	27.18	540	! S
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-722-187-6	-	49.00	119.97	27.18	540	! S
/cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:PCT-US95-04712-6	-	49.00	119.97	27.18	540	
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-096-762-199	+	52.50	117.88	35.53	1674	!
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-811-949-48	+	51.00	116.20	44.03	1314	!
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-811-949-50	+	51.00	116.20	44.03	1314	!
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-811-949-54	+	51.00	116.20	44.03	1314	!
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-811-949-56	+	51.00	116.20	44.03	1314	!
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-948-569A-7	-	52.00	116.09	44.66	1734	!
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-663-808-1	-	52.00	116.09	44.66	1734	!
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-615-961-9	+	39.00	115.71	46.93	56	! Se
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-811-949-62	+	51.00	115.39	48.88	1419	!
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-440-846-1	-	49.50	115.12	50.56	975	! S
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-885-970A-13	+	46.00	114.70	53.42	399	!
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-298-687A-13	+	46.00	114.70	53.42	399	!
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-530-797-12	-	46.00	114.70	53.42	399	!
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-298-829-13	+	46.00	114.70	53.42	399	!
/cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:PCT-US91-03012-4	-	44.00	114.60	54.05	236	
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-203-905B-4	-	49.00	114.20	56.92	931	!
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-737-524B-17	+	47.00	113.58	61.65	579	!
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-419-075-3	+	47.00	113.58	61.65	579	! S
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-791-347-9	-	42.50	112.66	69.33	190	! S
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-791-347-10	-	42.50	112.66	69.33	190	!
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-748-479-7	-	42.50	112.66	69.33	190	! S
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-748-479-8	-	42.50	112.66	69.33	190	! S
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-737-524B-12	+	47.00	112.54	70.47	639	!
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-419-075-2	+	47.00	112.54	70.47	639	! S
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-419-075-20	+	47.00	112.40	71.67	647	!
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-854-603-13	-	37.00	112.37	71.96	45	! S
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-270-805C-25	+	46.00	112.35	72.17	498	!
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-410-654B-25	+	46.00	112.35	72.17	498	!
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-474-851-25	+	46.00	112.35	72.17	498	!
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-481-560-25	+	46.00	112.35	72.17	498	!
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-419-075-18	+	47.00	112.31	72.58	653	!
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-811-949-38	+	51.00	111.90	76.50	1974	!
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-720-591A-1	+	47.50	111.32	82.35	819	!
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-811-949-42	+	51.00	111.24	83.25	2101	!
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-119-512-3	+	51.00	110.93	86.55	2162	!
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-488-015B-3	+	51.00	110.93	86.55	2162	
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-488-015B-25	+	51.00	110.93	86.55	2162	
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-101-506-1	+	38.00	110.67	89.48	69	!

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-347-221-1

seq_documentation_block:
; Sequence 1, Application US/08347221
; Patent No. 5525501
; GENERAL INFORMATION:
; APPLICANT: Masaharu TOKUYAMA et al.
; TITLE OF INVENTION: DNA FRAGMENT ENCODING ACYLAMINO ACID
; TITLE OF INVENTION: RACEMASE (AS AMENDED)
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 500 Kb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/347,221
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/984,310
; FILING DATE: December 1, 1992
; APPLICATION NUMBER: 07/668,475
; FILING DATE: March 13, 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX:
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1400 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE:
; HYPOTHETICAL:
; ANTI-SENSE:
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; ORGANISM:
; STRAIN:
; INDIVIDUAL ISOLATE:
; DEVELOPMENTAL STAGE:
; HAPLOTYPE:
; TISSUE TYPE:
; CELL TYPE:
; CELL LINE:
; ORGANELLE:
; IMMEDIATE SOURCE:
; LIBRARY:
; CLONE:
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT:
; MAP POSITION:
; UNITS:
; FEATURE:
; NAME/KEY:
; LOCATION:

```
;      IDENTIFICATION METHOD:
;      OTHER INFORMATION:
;    PUBLICATION INFORMATION:
;      AUTHORS:
;      TITLE:
;      JOURNAL:
;      VOLUME:
;      ISSUE:
;      PAGES:
;      DATE:
;      DOCUMENT NUMBER:
;      FILING DATE:
;      PUBLICATION DATE:
;      RELEVANT RESIDUES IN SEQ ID NO:
US-08-347-221-1


alignment_scores:
            Quality:   59.00            Length:     41
              Ratio:  1.788              Gaps:      1
 Percent Similarity:  80.488   Percent Identity: 31.707

alignment_block:
 US-09-068-804-8 x US-08-347-221-1   ..

 Align seg 1/1  to: US-08-347-221-1  from: 1  to: 1400

      13 MetProGlyMetGlnThrGluIleLysThrGlnAlaThrAsnLeuAlaAl 29
         |||:::   :::::::::|||:::::::::::::::::   ::::::|||||
     374 ATGGCCGTGCTCGACGCCGAACTCCGCGCGCACGAGAGGTCGTTCGCCGC 423

      29 aAsnLeuSerAlaValArgGluSerAlaThrAlaThrLeuSerGlyGluI 46
         |:::|||:::::::||||||:::|||::::::      :::|||      |
     424 CGAACTCGGATCGGTGCGCGATTCTGTGCCGTGCGGCGTTTCGGTCGGGA 473

      46 leLysGly......ProGlnLeu 51
         ||:::::::      |||||||||
     474 TCATGGACACCATCCCGCAACTG 496

seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-040-548-32

seq_documentation_block:
; Sequence 32, Application US/08040548
; Patent No. 5763209
;  GENERAL INFORMATION:
;    APPLICANT:  Sukhatme, Vikas P.
;    TITLE OF INVENTION:  METHODS AND MATERIALS RELATING TO THE
;    TITLE OF INVENTION:  FUNCTIONAL DOMAINS OF DNA BINDING PROTEINS
;    NUMBER OF SEQUENCES:  67
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  321 No. 5763209th Clark Street, Suite 800
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  U.S.A.
;      ZIP:  60610
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/040,548
;      FILING DATE:
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Coughlin, Daniel F.
;      REGISTRATION NUMBER:  36,111
;      REFERENCE/DOCKET NUMBER:  arcd067
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (312) 744-0090
;      TELEFAX:  (312) 245-4961
;  INFORMATION FOR SEQ ID NO:  32:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1200 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-040-548-32


alignment_scores:
            Quality:   52.00            Length:     45
              Ratio:  1.625              Gaps:      1
 Percent Similarity:  71.111   Percent Identity: 33.333

alignment_block:
 US-09-068-804-8 x US-08-040-548-32   ..

 Align seg 1/1  to: US-08-040-548-32  from: 1  to: 1200

       6 ArgThrGlnProProValIleMetProGlyMetGlnThrGluIleLysTh 22
         |||       ||||||          |||:::      ::::::      ::
     973 CGCGATTCGCCGCCGCCG......CCAGCTTCCGCCGCCGCAAGATCGGC 1016

      22 rGlnAlaThrAsnLeuAlaAlaAsnLeuSerAlaValArgGluSerAlaT 39
         ::::|||::::::   ||||||:::|||   :::   |||:::::::|||
    1017 CCCTGCCCCAGCCTCCGCGGCAGCCCTGCGTCCACCACGGGCCGCGGCTA 1066

      39 hrAlaThrLeuSerGlyGluIleLysGlyProGln 50
         |||||:::|||:::::::::::::   :::|||||
    1067 CCGCCAGCCTGGGGGCCCACCTACACTCCCCGCAG 1101

seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-466-344-32

seq_documentation_block:
; Sequence 32, Application US/08466344
; Patent No. 5773583
;  GENERAL INFORMATION:
;    APPLICANT:  Sukhatme, Vikas P.
;    TITLE OF INVENTION:  METHODS AND MATERIALS RELATING TO THE
;    TITLE OF INVENTION:  FUNCTIONAL DOMAINS OF DNA BINDING PROTEINS
;    NUMBER OF SEQUENCES:  67
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  321 No. 5773583th Clark Street, Suite 800
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  U.S.A.
;      ZIP:  60610
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/466,344
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/040,548
;      FILING DATE:  31-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Coughlin, Daniel F.
;      REGISTRATION NUMBER:  36,111
;      REFERENCE/DOCKET NUMBER:  arcd067
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (312) 744-0090
;      TELEFAX:  (312) 245-4961
;  INFORMATION FOR SEQ ID NO:  32:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1200 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
```

```
;    MOLECULE TYPE:  DNA (genomic)
US-08-466-344-32


alignment_scores:
            Quality:   52.00              Length:     45
              Ratio:  1.625                Gaps:      1
 Percent Similarity:  71.111   Percent Identity: 33.333

alignment_block:
 US-09-068-804-8 x US-08-466-344-32   ..

 Align seg 1/1  to: US-08-466-344-32  from: 1  to: 1200

       6 ArgThrGlnProProValIleMetProGlyMetGlnThrGluIleLysTh 22
         |||      ||||||         |||:::      ::::::      ::
     973 CGCGATTCGCCGCCGCCG......CCAGCTTCCGCCGCCGCAAGATCGGC 1016

      22 rGlnAlaThrAsnLeuAlaAlaAsnLeuSerAlaValArgGluSerAlaT 39
         ::::|||::::::   ||||||:::|||   :::   |||::::::|||| 
    1017 CCCTGCCCCAGCCTCCGCGGCAGCCCTGCGTCCACCACGGGCCGCGGCTA 1066

      39 hrAlaThrLeuSerGlyGluIleLysGlyProGln 50
         |||||::|||:::::::::::::   :::|||||| 
    1067 CCGCCAGCCTGGGGGCCCACCTACACTCCCCGCAG 1101

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-233-608-6

seq_documentation_block:
; Sequence 6, Application US/08233608
; Patent No. 5585238
;  GENERAL INFORMATION:
;    APPLICANT:  Ligon, James M
;    APPLICANT:  Beck, James J
;    TITLE OF INVENTION:  Detection of Fungal Pathogens Using the
;    TITLE OF INVENTION:  Polymerase Chain Reaction
;    NUMBER OF SEQUENCES:  49
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/233,608
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Spruill, W. Murray
;      REGISTRATION NUMBER:  32,943
;      REFERENCE/DOCKET NUMBER:  CGC 1739
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8615
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  540 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Mycosphaerella musicola
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..540
;      OTHER INFORMATION:  /note= "DNA sequence for the
;      OTHER INFORMATION:  Internal Transcribed Spacer of Mycosphaerella
;      OTHER INFORMATION:  musicola"
US-08-233-608-6


alignment_scores:
            Quality:   49.00              Length:     45
              Ratio:  1.531                Gaps:      1
 Percent Similarity:  71.111   Percent Identity: 26.667

alignment_block:
 US-09-068-804-8 x US-08-233-608-6/rev   ..

 Align seg 1/1  to reverse of: US-08-233-608-6  from: 1  to: 540

       6 ArgThrGlnProProValIleMetProGlyMetGlnThrGluIleLysTh 22
         |||::::::||||||::::::   :::   :::::::::      :::
     467 CGGCCGCGGCCACCCGCGCTCCGGAGCGAACTGAAAAATCCACAACGCTT 418

      22 rGlnAlaThrAsnLeuAlaAlaAsnLeuSerAlaValArgGluSerAlaT 39
         ::::::   :::|||||:::   |||   :::::::|||   :::|||: 
     417 AGAGACGGACGGCTTAGCCGGGAGACTTTGGGGCGCGCGGAGCACCGCGG 368

      39 hrAlaThr.............LeuSerGlyGluIle 46
         ::::::::             ||||||||||||::: 
     367 CGCCCAATACCAAGCTAGGCTTGAGTGGTGAAATG 333

seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-887-480-6

seq_documentation_block:
; Sequence 6, Application US/08887480
; Patent No. 5814453
;  GENERAL INFORMATION:
;    APPLICANT:  Beck, James J
;    TITLE OF INVENTION:  Detection of Fungal Pathogens Using the
;    TITLE OF INVENTION:  Polymerase Chain Reaction
;    NUMBER OF SEQUENCES:  96
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5814453artis Corporation
;      STREET:  520 White Plains Road
;      CITY:  Tarrytown
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/887,480
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/722,187
;      FILING DATE:  15-OCT-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  CGC 1739/PCT/CIP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  540 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
```

```
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Mycosphaerella musicola
;    FEATURE:
;      NAME/KEY:  misc_ feature
;      LOCATION:  1..30
;      OTHER INFORMATION:  /note= "3' end of small subunit
;      OTHER INFORMATION:  rRNA gene"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  31..180
;      OTHER INFORMATION:  /note= "ITS 1"
;    FEATURE:
;      NAME/KEY:  misc_ feature
;      LOCATION:  181..336
;      OTHER INFORMATION:  /note= "5.8S rRNA gene"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  337..483
;      OTHER INFORMATION:  /note= "ITS 2"
;    FEATURE:
;      NAME/KEY:  misc_ feature
;      LOCATION:  484..540
;      OTHER INFORMATION:  /note= "5' end of large subunit
;      OTHER INFORMATION:  rRNA gene"
US-08-887-480-6


alignment_scores:
            Quality:   49.00              Length:     45
              Ratio:  1.531                Gaps:      1
 Percent Similarity:  71.111   Percent Identity: 26.667

alignment_block:
 US-09-068-804-8 x US-08-887-480-6/rev   ..

 Align seg 1/1  to reverse of: US-08-887-480-6  from: 1  to: 540

       6 ArgThrGlnProProValIleMetProGlyMetGlnThrGluIleLysTh 22
         |||:::::::||||||:::::::   :::   :::::::::      :::
     467 CGGCCGCGGCCACCCGCGCTCCGGAGCGAACTGAAAAATCCACAACGCTT 418

      22 rGlnAlaThrAsnLeuAlaAlaAsnLeuSerAlaValArgGluSerAlaT 39
         ::::::   :::|||||||:::   |||   ::::::|||   :::|||:
     417 AGAGACGGACGGCTTAGCCGGGAGACTTTGGGGCGCGCGGAGCACCGCGG 368

      39 hrAlaThr............LeuSerGlyGluIle 46
         ::::::::            ||||||||||||:::
     367 CGCCCAATACCAAGCTAGGCTTGAGTGGTGAAATG 333

seq_name: /cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-722-187-6

seq_documentation_block:
; Sequence 6, Application US/08722187
; Patent No. 5955274
;  GENERAL INFORMATION:
;    APPLICANT:  Ligon, James M
;    APPLICANT:  Beck, James J
;    TITLE OF INVENTION:  Detection of Fungal Pathogens Using the
;    TITLE OF INVENTION:  Polymerase Chain Reaction
;    NUMBER OF SEQUENCES:  86
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/722,187
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/233,608
;      FILING DATE:  04-APR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Walsh, Andrea C.
;      REGISTRATION NUMBER:  34,988
;      REFERENCE/DOCKET NUMBER:  CGC 1739
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8666
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  540 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Mycosphaerella musicola
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..540
;      OTHER INFORMATION:  /note= "DNA sequence for the
;      OTHER INFORMATION:  Internal Transcribed Spacer of Mycosphaerella
;      OTHER INFORMATION:  musicola"
US-08-722-187-6


alignment_scores:
            Quality:   49.00              Length:     45
              Ratio:  1.531                Gaps:      1
 Percent Similarity:  71.111   Percent Identity: 26.667

alignment_block:
 US-09-068-804-8 x US-08-722-187-6/rev   ..

 Align seg 1/1  to reverse of: US-08-722-187-6  from: 1  to: 540

       6 ArgThrGlnProProValIleMetProGlyMetGlnThrGluIleLysTh 22
         |||:::::::||||||:::::::   :::   :::::::::      :::
     467 CGGCCGCGGCCACCCGCGCTCCGGAGCGAACTGAAAAATCCACAACGCTT 418

      22 rGlnAlaThrAsnLeuAlaAlaAsnLeuSerAlaValArgGluSerAlaT 39
         ::::::   :::|||||||:::   |||   ::::::|||   :::|||:
     417 AGAGACGGACGGCTTAGCCGGGAGACTTTGGGGCGCGCGGAGCACCGCGG 368

      39 hrAlaThr............LeuSerGlyGluIle 46
         ::::::::            ||||||||||||:::
     367 CGCCCAATACCAAGCTAGGCTTGAGTGGTGAAATG 333

seq_name: /cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:PCT-US95-04712-6

seq_documentation_block:
; Sequence 6, Application PC/TUS9504712
;  GENERAL INFORMATION:
;    APPLICANT:  Ligon, James M
;    APPLICANT:  Beck, James J
;    TITLE OF INVENTION:  Detection of Fungal Pathogens Using the
;    TITLE OF INVENTION:  Polymerase Chain Reaction
;    NUMBER OF SEQUENCES:  86
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
```

```
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US95/04712
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/233,608
;       FILING DATE:  04-APR-1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Walsh, Andrea C.
;       REGISTRATION NUMBER:  34,988
;       REFERENCE/DOCKET NUMBER:  CGC 1739
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  919-541-8666
;       TELEFAX:  919-541-8689
;   INFORMATION FOR SEQ ID NO:  6:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  540 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  Mycosphaerella musicola
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  1..540
;       OTHER INFORMATION:  /note= "DNA sequence for the
;       OTHER INFORMATION:  Internal Transcribed Spacer of Mycosphaerella
;       OTHER INFORMATION:  musicola"
PCT-US95-04712-6


alignment_scores:
            Quality:   49.00            Length:     45
              Ratio:  1.531              Gaps:      1
 Percent Similarity:  71.111   Percent Identity: 26.667

alignment_block:
 US-09-068-804-8 x PCT-US95-04712-6/rev   ..

 Align seg 1/1  to reverse of: PCT-US95-04712-6  from: 1  to: 540

       6 ArgThrGlnProProValIleMetProGlyMetGlnThrGluIleLysTh 22
         |||:::::::||||||:::::   :::   :::::::::     :::
     467 CGGCCGCGGCCACCCGCGCTCCGGAGCGAACTGAAAAATCCACAACGCTT 418

      22 rGlnAlaThrAsnLeuAlaAlaAsnLeuSerAlaValArgGluSerAlaT 39
         ::::::   :::|||||||:::   |||   ::::::|||   :::|||:
     417 AGAGACGGACGGCTTAGCCGGGAGACTTTGGGGCGCGCGGAGCACCGCGG 368

      39 hrAlaThr.............LeuSerGlyGluIle 46
         ::::::::             ||||||||||||:::
     367 CGCCCAATACCAAGCTAGGCTTGAGTGGTGAAATG 333

seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-096-762-199

seq_documentation_block:
; Sequence 199, Application US/08096762
; Patent No. 5814318
;  GENERAL INFORMATION:
;     APPLICANT:  Lonberg, Nils
;     APPLICANT:  Kay, Robert M.
;     TITLE OF INVENTION:  Transgenic No. 5814318-Human Animals for
;     TITLE OF INVENTION:  Producing Heterologous Antibodies
;     NUMBER OF SEQUENCES:  210
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Townsend and Townsend Khourie and Crew
;       STREET:  One Market Plaza, Steuart Tower, Suite 200
;       CITY:  San Francisco
;       STATE:  California
;       COUNTRY:  USA
;       ZIP:  94105
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/096,762
;       FILING DATE:  22-JUL-1993
;       CLASSIFICATION:  800
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/053,131
;       FILING DATE:  26-APR-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/990,860
;       FILING DATE:  16-DEC-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/904,068
;       FILING DATE:  23-JUN-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/853,408
;       FILING DATE:  18-MAR-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/810,279
;       FILING DATE:  17-DEC-1991
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Smith, William M.
;       REGISTRATION NUMBER:  30,223
;       REFERENCE/DOCKET NUMBER:  14643-9-4
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  415-326-2400
;       TELEFAX:  415-326-2422
;   INFORMATION FOR SEQ ID NO:  199:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1674 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:  NO
US-08-096-762-199


alignment_scores:
            Quality:   52.50            Length:     53
              Ratio:  1.382              Gaps:      1
 Percent Similarity:  71.698   Percent Identity: 30.189

alignment_block:
 US-09-068-804-8 x US-08-096-762-199   ..

 Align seg 1/1  to: US-08-096-762-199  from: 1  to: 1674

       2 ValThrSerValArgThrGlnProProValIleMetProGlyMetGlnTh 18
         ||||||:::      :::::|||||||:::::::::|||:::   |||::
     582 GTTACAGGTCCTACTCCTCCTCCTCCTATTACTATTCCTTCCTGCCAGCC 631

      18 rGluIleLysThrGlnAlaThrAsnLeuAlaAlaAsnLeuSerAlaValA 35
         :   :::      |||   ::::::|||:::::::   |||   :::
     632 CAGCCTGTCACTGCAGCGGCCAGCTCTTGAGGACCTGCTCCTGGGTTCAG 681

      35 rgGluSerAlaThrAlaThrLeuSerGlyGluIleLysGlyProGlnLeu 51
           :::|||   |||   |||||:::|||   ::::::::::|||:::
     682 ATGCCAGCATCACATGTACTCTGAATGGC...CTGAGAAATCCTGAGGGA 728

      52 GluAspPhe 54
         :::::|||
     729 GCTGCTTTC 737
```

```
seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-811-949-48

seq_documentation_block:
; Sequence 48, Application US/08811949
; Patent No. 5840533
;  GENERAL INFORMATION:
;    APPLICANT:  NIWA, MINEO
;    APPLICANT:  SAITO, YOSHIMASA
;    APPLICANT:  SASAKI, HITOSHI
;    APPLICANT:  HAYASHI, MASAKO
;    APPLICANT:  NOTANI, JOUJI
;    APPLICANT:  KOBAYASHI, MASAKAZU
;    TITLE OF INVENTION:  TISSUE PLASMINOGEN ACTIVATOR
;    NUMBER OF SEQUENCES:  67
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
;      CITY:  ARLINGTON
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/811,949
;      FILING DATE:  05-MAR-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OBLON, NORMAN F.
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  18-966-0
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-413-3000
;      TELEFAX:  703-413-2220
;  INFORMATION FOR SEQ ID NO:  48:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1314 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  circular
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1311
US-08-811-949-48


alignment_scores:
           Quality:   51.00            Length:     39
             Ratio:  2.040              Gaps:      2
 Percent Similarity:  64.103   Percent Identity: 43.590

alignment_block:
 US-09-068-804-8 x US-08-811-949-48   ..

 Align seg 1/1  to: US-08-811-949-48  from: 1  to: 1314

      17 GlnThrGluIleLysThrGln.........AlaThrAsnLeu........ 27
         :::|||:::|||   ||||||         ||||||:::|||
     164 GAAACCCAGATCGAGACTCAAAGCCCTGGTGCTACGTCTTTAAGGCGGGG 213

      28 ...................AlaAlaAsnLeuSerAlaValArgGluSerA 38
                             ||||||   |||:::|||:::||||||:::
     214 AAGTACAGCTCAGAGTTCTGCAGCACCCCTGCCTGCTCTGAGGGAAACAG 263

     38 laThrAlaThrLeuSer 43
        ::|||||||||||||:::
     264 TGACTGCTACTTTGGGA 280
```

```
seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-811-949-50

seq_documentation_block:
; Sequence 50, Application US/08811949
; Patent No. 5840533
;  GENERAL INFORMATION:
;    APPLICANT:  NIWA, MINEO
;    APPLICANT:  SAITO, YOSHIMASA
;    APPLICANT:  SASAKI, HITOSHI
;    APPLICANT:  HAYASHI, MASAKO
;    APPLICANT:  NOTANI, JOUJI
;    APPLICANT:  KOBAYASHI, MASAKAZU
;    TITLE OF INVENTION:  TISSUE PLASMINOGEN ACTIVATOR
;    NUMBER OF SEQUENCES:  67
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
;      CITY:  ARLINGTON
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/811,949
;      FILING DATE:  05-MAR-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OBLON, NORMAN F.
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  18-966-0
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-413-3000
;      TELEFAX:  703-413-2220
;  INFORMATION FOR SEQ ID NO:  50:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1314 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  circular
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1311
US-08-811-949-50


alignment_scores:
           Quality:   51.00            Length:     39
             Ratio:  2.040              Gaps:      2
 Percent Similarity:  64.103   Percent Identity: 43.590

alignment_block:
 US-09-068-804-8 x US-08-811-949-50   ..

 Align seg 1/1  to: US-08-811-949-50  from: 1  to: 1314

      17 GlnThrGluIleLysThrGln.........AlaThrAsnLeu........ 27
         :::|||:::|||   ||||||         ||||||:::|||
     164 GAAACCCAGATCGAGACTCAAAGCCCTGGTGCTACGTCTTTAAGGCGGGG 213

      28 ...................AlaAlaAsnLeuSerAlaValArgGluSerA 38
                             ||||||   |||:::|||:::||||||:::
     214 AAGTACAGCTCAGAGTTCTGCAGCACCCCTGCCTGCTCTGAGGGAAACAG 263

      38 laThrAlaThrLeuSer 43
         ::|||||||||||||:::
     264 TGACTGCTACTTTGGGA 280
```

```
seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-811-949-54

seq_documentation_block:
; Sequence 54, Application US/08811949
; Patent No. 5840533
;  GENERAL INFORMATION:
;    APPLICANT:  NIWA, MINEO
;    APPLICANT:  SAITO, YOSHIMASA
;    APPLICANT:  SASAKI, HITOSHI
;    APPLICANT:  HAYASHI, MASAKO
;    APPLICANT:  NOTANI, JOUJI
;    APPLICANT:  KOBAYASHI, MASAKAZU
;    TITLE OF INVENTION:  TISSUE PLASMINOGEN ACTIVATOR
;    NUMBER OF SEQUENCES:  67
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
;      CITY:  ARLINGTON
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/811,949
;      FILING DATE:  05-MAR-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OBLON, NORMAN F.
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  18-966-0
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-413-3000
;      TELEFAX:  703-413-2220
;  INFORMATION FOR SEQ ID NO:  54:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1314 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  circular
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1311
US-08-811-949-54


alignment_scores:
            Quality:   51.00              Length:     39
              Ratio:  2.040                Gaps:      2
 Percent Similarity:  64.103   Percent Identity: 43.590

alignment_block:
 US-09-068-804-8 x US-08-811-949-54    ..

 Align seg 1/1  to: US-08-811-949-54  from: 1  to: 1314

      17 GlnThrGluIleLysThrGln.........AlaThrAsnLeu........ 27
         :::|||:::|||   ||||||         ||||||:::|||
     164 GAAACCCAGATCGAGACTCAAAGCCCTGGTGCTACGTCTTTAAGGCGGGG 213

      28 ...................AlaAlaAsnLeuSerAlaValArgGluSerA 38
                             ||||||   |||:::|||:::||||||:::
     214 AAGTACAGCTCAGAGTTCTGCAGCACCCCTGCCTGCTCTGAGGGAAACAG 263

      38 laThrAlaThrLeuSer 43
         ::|||||||||||:::
     264 TGACTGCTACTTTGGGA 280
```

```
seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-811-949-56

seq_documentation_block:
; Sequence 56, Application US/08811949
; Patent No. 5840533
;  GENERAL INFORMATION:
;    APPLICANT:  NIWA, MINEO
;    APPLICANT:  SAITO, YOSHIMASA
;    APPLICANT:  SASAKI, HITOSHI
;    APPLICANT:  HAYASHI, MASAKO
;    APPLICANT:  NOTANI, JOUJI
;    APPLICANT:  KOBAYASHI, MASAKAZU
;    TITLE OF INVENTION:  TISSUE PLASMINOGEN ACTIVATOR
;    NUMBER OF SEQUENCES:  67
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
;      CITY:  ARLINGTON
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/811,949
;      FILING DATE:  05-MAR-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OBLON, NORMAN F.
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  18-966-0
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-413-3000
;      TELEFAX:  703-413-2220
;  INFORMATION FOR SEQ ID NO:  56:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1314 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  circular
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1311
US-08-811-949-56


alignment_scores:
            Quality:   51.00              Length:     39
              Ratio:  2.040                Gaps:      2
 Percent Similarity:  64.103   Percent Identity: 43.590

alignment_block:
 US-09-068-804-8 x US-08-811-949-56    ..

 Align seg 1/1  to: US-08-811-949-56  from: 1  to: 1314

      17 GlnThrGluIleLysThrGln.........AlaThrAsnLeu........ 27
         :::|||:::|||   ||||||         ||||||:::|||
     164 GAAACCCAGATCGAGACTCAAAGCCCTGGTGCTACGTCTTTAAGGCGGGG 213

      28 ...................AlaAlaAsnLeuSerAlaValArgGluSerA 38
                             ||||||   |||:::|||:::||||||:::
     214 AAGTACAGCTCAGAGTTCTGCAGCACCCCTGCCTGCTCTGAGGGAAACAG 263

      38 laThrAlaThrLeuSer 43
         ::|||||||||||:::
     264 TGACTGCTACTTTGGGA 280
```

```
seq_name: /cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-948-569A-7

seq_documentation_block:
; Sequence 7, Application US/08948569A
; Patent No. 5882926
;  GENERAL INFORMATION:
;    APPLICANT:  Amara, Susan G
;    APPLICANT:  Arriza, Jeffrey L
;    APPLICANT:  Eliasof, Scott
;    APPLICANT:  Kavanaugh, Michael P
;    TITLE OF INVENTION:  Excitatory Amino Acid Transporter Genes
;    TITLE OF INVENTION:  and Uses
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  McDonnell Boehnen Hulbert & Berghoff
;      STREET:  300 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/948,569A
;      FILING DATE:  10-OCT-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 5882926nan, Kevin E
;      REGISTRATION NUMBER:  35,303
;      REFERENCE/DOCKET NUMBER:  93,509-F
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312-913-0001
;      TELEFAX:  312-913-0002
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1734 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  9..1700
US-08-948-569A-7


alignment_scores:
            Quality:   52.00              Length:     41
              Ratio:  1.857                Gaps:      0
 Percent Similarity:  68.293   Percent Identity: 34.146

alignment_block:
 US-09-068-804-8 x US-08-948-569A-7/rev   ..

 Align seg 1/1  to reverse of: US-08-948-569A-7  from: 1  to: 1734

       4 SerValArgThrGlnProProValIleMetProGlyMetGlnThrGluIl 20
         |||:::|||||   |||      :::||||||   |||   :::
     889 TCCCTGAGGACTCTGCCCTTGTGTTTCATGCCACCAATGACCAGCCCAAA 840

      20 eLysThrGlnAlaThrAsnLeuAlaAlaAsnLeuSerAlaValArgGluS 37
            |||:::   |||:::   :::   :::|||   ::::::   |||:
     839 GGCCACAGAGAAGACCACGAGGCCCAGGGCGTTGATGCCATTGGCGGAGC 790

      37 erAlaThrAlaThrLeuSerGly 44
         :::::|||:::|||:::|||:::
     789 CAGGCACGGGTACAGTCTCCTCA 767

seq_name: /cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-663-808-1

seq_documentation_block:
; Sequence 1, Application US/08663808
; Patent No. 5912171
;  GENERAL INFORMATION:
;    APPLICANT:  Amara, Susan G
;    APPLICANT:  Arriza, Jeffrey L
;    APPLICANT:  Fairman, Wendy A
;    TITLE OF INVENTION:  Amino Acid Transporters and Uses
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Banner & Allegretti, Ltd.
;      STREET:  10 South Wacker Drive, Suite 3000
;      CITY:  Chicago
;      STATE:  IL
;      COUNTRY:  USA
;      ZIP:  60606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/663,808
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 5912171nan, Kevin E
;      REGISTRATION NUMBER:  35,303
;      REFERENCE/DOCKET NUMBER:  93,509-C
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312-715-1000
;      TELEFAX:  312-715-1234
;      TELEX:  910-221-5317
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1734 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  9..1703
;    FEATURE:
;      NAME/KEY:  5'UTR
;      LOCATION:  1..8
;    FEATURE:
;      NAME/KEY:  3'UTR
;      LOCATION:  1704..1734
US-08-663-808-1


alignment_scores:
            Quality:   52.00              Length:     41
              Ratio:  1.857                Gaps:      0
 Percent Similarity:  68.293   Percent Identity: 34.146

alignment_block:
 US-09-068-804-8 x US-08-663-808-1/rev   ..

 Align seg 1/1  to reverse of: US-08-663-808-1  from: 1  to: 1734

       4 SerValArgThrGlnProProValIleMetProGlyMetGlnThrGluIl 20
         |||:::|||||   |||      :::||||||   |||   :::
     889 TCCCTGAGGACTCTGCCCTTGTGTTTCATGCCACCAATGACCAGCCCAAA 840

      20 eLysThrGlnAlaThrAsnLeuAlaAlaAsnLeuSerAlaValArgGluS 37
            |||:::   |||:::   :::   :::|||   ::::::   |||:
     839 GGCCACAGAGAAGACCACGAGGCCCAGGGCGTTGATGCCATTGGCGGAGC 790

      37 erAlaThrAlaThrLeuSerGly 44
         :::::|||:::|||:::|||:::
```

789 CAGGCACGGGTACAGTCTCCTCA 767

seq_name: /cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-615-961-9

```
seq_documentation_block:
; Sequence 9, Application US/08615961
; Patent No. 5877162
;  GENERAL INFORMATION:
;    APPLICANT:  Martina Werner and Shaji T. George
;    TITLE OF INVENTION:  Short External Guide Sequences
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Patrea L. Pabst
;      STREET:  2800 One Atlantic Center
;      STREET:  1201 West Peachtree Street
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  USA
;      ZIP:  30309-3450
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/615,961
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pabst, Patrea L.
;      REGISTRATION NUMBER:  31,284
;      REFERENCE/DOCKET NUMBER:  ILI115
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (404) 873-8794
;      TELEFAX:  (404) 873-8795
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  56 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  RNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-615-961-9


alignment_scores:
            Quality:   39.00              Length:     12
              Ratio:  3.545                Gaps:      0
 Percent Similarity:  91.667   Percent Identity: 58.333

alignment_block:
 US-09-068-804-8 x US-08-615-961-9    ..

 Align seg 1/1  to: US-08-615-961-9  from: 1  to: 56

       9 ProProValIleMetProGlyMetGlnThrGluIle 20
         ||||||||||   :::|||:::::::::|||||||||
      18 CCUCCUGUUGCUGUACCAAACCUUCGGACGGAAAUU 53
```

This Page Blank (uspto)

OM of: US-09-068-804-8 to: N_Geneseq_36:* out_format : pfs

Date: Oct 5, 1999 8:04 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09068804/runat_05101999_145135_10341/app_query.fasta.1
-DB=N_Geneseq_36 -QFMT=fastap -SUFFIX=rng -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=pam150
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=escore
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09068804
-NCPU=6 -ICPU=3 -WAIT -THREADS=1

Search information block:
Query: US-09-068-804-8
Query length: 64
Database: N_Geneseq_36:*
Database sequences: 311585
Database length: 125096042
Search time (sec): 156.120000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ..
N_Geneseq_36:T67031	+	277.00	743.57	8.1e-34	194	! Salmonella secreted protein Ssp
N_Geneseq_36:T67035	+	264.00	678.81	3.3e-30	3441	! Salmonella virulence factor Ss
N_Geneseq_36:X13094	-	60.50	141.38	2.83	1014	! Enterococcus faecalis genome c
N_Geneseq_36:Q22689	+	59.00	134.00	7.30	1400	! Sequence encoding acylamino ac
N_Geneseq_36:V77848	+	54.00	132.79	8.53	424	! Staphylococcus aureus contig SE
N_Geneseq_36:V77883	+	52.00	127.98	15.79	400	! Staphylococcus aureus contig SE
N_Geneseq_36:N92320	+	48.00	123.38	28.51	219	! Murine autonomously replicating
N_Geneseq_36:X00509	+	52.00	122.16	33.33	704	! Rat EGR-1 gene promoter. Replic
N_Geneseq_36:Q29170	-	45.00	121.73	35.23	117	! HLA-Bw 52 exon 5. Transgenic nc
N_Geneseq_36:V89595	+	51.00	121.10	38.16	600	! EST clone CP147. New polynuclec
N_Geneseq_36:Q12072	+	55.50	120.36	41.97	2100	! T-PA with -ve charged finger a
N_Geneseq_36:T98639	+	54.00	119.32	47.99	1568	! DNA encoding a S. pneumoniae r
N_Geneseq_36:T22145	-	48.00	117.73	58.84	379	! Human gene signature HUMGS03715
N_Geneseq_36:V77996	+	48.00	117.17	63.19	400	! Staphylococcus aureus contig SE
N_Geneseq_36:T05399	-	49.00	116.78	66.42	540	! Mycosphaerella musicola interna
N_Geneseq_36:V62506	-	49.00	116.78	66.42	540	! Mycosphaerella musicola PCR amp
N_Geneseq_36:N91817	+	52.00	116.67	67.42	1200	! Regulatory region of Egr-1-con
N_Geneseq_36:Q73465	+	52.00	116.67	67.42	1200	! Mouse Egr-1 genomic clone. New
N_Geneseq_36:T80769	+	50.50	116.65	67.58	811	! Staphylococcus aureus Gene #8 e
N_Geneseq_36:V07910	+	51.00	116.61	67.90	928	! Human cysteine-rich secreted pr
N_Geneseq_36:X34080	+	49.00	116.50	68.86	555	! Mycobacterium species nucleic a
N_Geneseq_36:N80896	+	51.00	116.35	70.23	952	! Recombinant human tissue plasmi
N_Geneseq_36:X40441	+	46.50	115.32	80.13	323	! Human secreted protein 5' EST s
N_Geneseq_36:T30898	-	42.00	115.24	81.00	100	! Primer 28 for 95 kD protein cod
N_Geneseq_36:X34081	+	49.00	115.10	82.44	636	! Mycobacterium species nucleic a
N_Geneseq_36:X18059	-	44.00	114.76	86.10	177	! Coding sequence for human SI bi
N_Geneseq_36:V12732	+	52.50	114.59	88.01	1674	! Mouse heavy chain locus alpha
N_Geneseq_36:V38199	+	52.50	114.59	88.01	1674	! Mouse heavy chain locus alpha
N_Geneseq_36:X37434	+	49.50	114.19	92.63	792	! Human secreted protein cDNA fra
N_Geneseq_36:V90357	-	45.00	114.16	93.04	244	! EST clone DL320. New polynuclec
N_Geneseq_36:Q87912	-	53.00	113.96	95.41	2029	! Glutathione-reductase cDNA. Fu
N_Geneseq_36:X51810	-	47.00	113.75	98.02	429	! Human secreted protein 5' EST s
N_Geneseq_36:V30384	-	39.00	113.67	99.06	53	! Oligomer p53rg22 used in constru
N_Geneseq_36:T85473	+	57.50	113.66	99.15	6803	! Genomic hNET sequence from P1
N_Geneseq_36:V16305	+	57.50	113.66	99.15	6803	! Genomic DNA encoding human net
N_Geneseq_36:Q39876	-	45.50	113.62	99.63	293	! Expressed Sequence Tag human ge
N_Geneseq_36:Q59288	-	45.50	113.62	99.63	293	! Human brain Expressed Sequence
N_Geneseq_36:N80141	+	44.50	113.50	101.16	228	! Synthetic t-PA kringle domain #
N_Geneseq_36:Q81552	+	44.50	113.50	101.16	228	! Kringle 2 domain fragment for c
N_Geneseq_36:V87854	+	46.50	113.27	104.18	394	! EST clone FC12. New polynucleot
N_Geneseq_36:T91633	+	39.00	113.10	106.53	56	! Short external guide sequence ta
N_Geneseq_36:N91117	+	51.00	113.03	107.50	1314	! Sequence of coding region in p
N_Geneseq_36:N91123	+	51.00	113.03	107.50	1314	! Sequence of coding region i
N_Geneseq_36:N91126	+	51.00	113.03	107.50	1314	! Sequence of coding region i
N_Geneseq_36:N91127	+	51.00	113.03	107.50	1314	! Sequence of coding region i

seq_name: N_Geneseq_36:T67031

seq_documentation_block:
ID T67031 standard; DNA; 194 BP.
AC T67031;
DT 04-AUG-1997 (first entry)
DE Salmonella secreted protein SspA DNA.
KW Salmonella secreted protein; SspA; bacterial-mediated endocytosis;
KW diagnosis; therapy; vaccine; attenuation; virulence; ss.
OS Salmonella typhimurium.
FH Key Location/Qualifiers
FT cds 1. .194
FT /*tag= a
FT /product= SspA (incomplete)
FT /note= "the translated sequence does not fully
FT correspond to the SspA amino acid sequence
FT given in Fig 23"
PN WO9718225-A1.
PD 22-MAY-1997.
PF 14-NOV-1996; U18504.
PR 14-NOV-1995; US-006733.
PA (GEHO) GEN HOSPITAL CORP.
PI Miller SI;
DR WPI; 97-289217/26.
DR P-PSDB; W15271.
PT New isolated Salmonella secreted proteins and related genes - used
PT to develop products for the detection, treatment or prevention of
PT Salmonella infections
PS Claim 9; Fig 22; 95pp; English.
CC A DNA molecule (T67031) codes for (incomplete) Salmonella secreted
CC protein SspA (W15271), a virulence factor able to translocate into
CC human epithelial cells. To identify Ssp genes, S. typhimurium
CC mutants were examined in invasion assays on HEp-2 epithelial cells.
CC 2 Ssp mutants with transposon insertions located between spaT and
CC prgH were identified. Of these, one mutant lacked Ssp of 87,
CC 42 and 36 kDa. The invasiveness of this mutant implicated the
CC 42 and 36 kDa Ssp in invasion. DNA sequencing of this region (see
CC also T67035) identified 2 complete and 2 partial open reading
CC frames, designated sspB, sspC (T67029), sspD (T67030) and sspA
CC (T67031). These nucleic acids can be used in methods for diagnosing
CC Salmonella infection and in methods of targetting an antigen to
CC epithelial cells and of inducing a cytotoxic T cell immune response
CC in a mammal. Ssp gene mutation can be used to create attenuated
CC Salmonella strains useful as live vaccines.
SQ Sequence 194 BP; 55 A; 51 C; 53 G; 35 T;

alignment_scores:
Quality: 277.00 Length: 64
Ratio: 4.328 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-068-804-8 x T67031 ..

Align seg 1/1 to: T67031 from: 1 to: 194

```
      1 MetValThrSerValArgThrGlnProProValIleMetProGlyMetGl 17
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      1 ATGGTTACAAGTGTAAGGACTCAGCCCCCCGTCATAATGCCAGGTATGCA 50

     17 nThrGluIleLysThrGlnAlaThrAsnLeuAlaAlaAsnLeuSerAlaV 34
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     51 GACCGAGATCAAAACGCAGGCCACGAATCTTGCGGCGAATCTTTCCGCAG 100

     34 alArgGluSerAlaThrAlaThrLeuSerGlyGluIleLysGlyProGln 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    101 TCAGAGAAAGTGCCACAGCGACGCTGTCAGGGGAAATTAAAGGCCCGCAA 150
```

```
     51 LeuGluAspPheProAlaLeuIleLysGlnAlaSerLeuAsp 64
        ||||||||||||||||||||||||||||||||||||||||||||
    151 CTGGAAGATTTTCCCGCGCTGATCAAACAGGCGAGTCTGGAT 192
```

seq_name: N_Geneseq_36:T67035

seq_documentation_block:
ID T67035 standard; DNA; 3441 BP.
AC T67035;
DT 04-AUG-1997 (first entry)
DE Salmonella virulence factor SspBCDA DNA region.
KW Salmonella secreted protein; SspB; SspC; SspD; SspA;
KW bacterial-mediated endocytosis; diagnosis; therapy; vaccine;
KW attenuation; virulence; ss.
OS Salmonella typhimurium.

FH	Key	Location/Qualifiers
FT	cds	1. .870
FT		/*tag= a
FT		/product= SspB (incomplete)
FT	cds	898. .2127
FT		/*tag= b
FT		/product= SspC
FT	cds	2198. .3228
FT		/*tag= c
FT		/product= SspD
FT	cds	3247. .3441
FT		/*tag= d
FT		/product= SspA (incomplete)

PN WO9718225-A1.
PD 22-MAY-1997.
PF 14-NOV-1996; U18504.
PR 14-NOV-1995; US-006733.
PA (GEHO) GEN HOSPITAL CORP.
PI Miller SI;
DR WPI; 97-289217/26.
DR P-PSDB; W15268-71.
PT New isolated Salmonella secreted proteins and related genes - used
PT to develop products for the detection, treatment or prevention of
PT Salmonella infections
PS Claim 11; Fig 28; 95pp; English.
CC A DNA molecule (T67035) comprises the SspBCDA locus (truncated at
CC the 5' and 3' ends) including the partial sspB, sspC, sspD and
CC partial sspA sequences (see also T67028-31) coding for Salmonella
CC typhimurium secreted proteins SspB, SspC, SspD and SspA (W15268-
CC 70). To identify the ssp genes, S. typhimurium mutants were
CC investigated in invasion assays on HEp-2 epithelial cells. 2
CC Ssp mutants with transposon insertions located between spaT and
CC prgH (see also T67036) were identified. One mutant lacked Ssp of
CC 87, 42 and 36 kDa. The invasiveness of this mutant implicated the
CC 42 and 36 kDa Ssp in invasion. DNA sequencing of this region
CC identified the 4 ssp (partial) genes. The ssp genes can be used in
CC methods for diagnosing Salmonella infection and in methods of
CC targetting an antigen to epithelial cells and of inducing a
CC cytotoxic T cell immune response in a mammal. Ssp gene mutation
CC can be used to create attenuated Salmonella strains useful as live
CC vaccines.
SQ Sequence 3441 BP; 978 A; 761 C; 925 G; 777 T;

alignment_scores:
Quality: 264.00 Length: 65
Ratio: 4.125 Gaps: 1
Percent Similarity: 98.462 Percent Identity: 98.462

alignment_block:
US-09-068-804-8 x T67035 ..

Align seg 1/1 to: T67035 from: 1 to: 3441

```
      1 MetValThrSerValArgThrGln.ProProValIleMetProGlyMetG 17
        ||||||||||||||||||||||| ||||||||||||||||||||||||||
   3317 ATGGTTACAAGTGTAAGGACTCAGCCCCCCCGTCATAATGCCAGGTATGC 3296
```

```
     17 lnThrGluIleLysThrGlnAlaThrAsnLeuAlaAlaAsnLeuSerAla 33
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   3297 AGACCGAGATCAAAACGCAGGCCACGAATCTTGCGGCGAATCTTTCCGCA 3346

     34 ValArgGluSerAlaThrAlaThrLeuSerGlyGluIleLysGlyProGl 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   3347 GTCAGAGAAAGTGCCACAGCGACGCTGTCAGGGGAAATTAAAGGCCCGCA 3396

     50 nLeuGluAspPheProAlaLeuIleLysGlnAlaSerLeuAsp 64
        |||||||||||||||||||||||||||||||||||||||||||
   3397 ACTGGAAGATTTTCCCGCGCTGATCAAACAGGCGAGTCTGGAT 3439
```

seq_name: N_Geneseq_36:X13094

seq_documentation_block:
ID X13094 standard; DNA; 1014 BP.
AC X13094;
DT 19-MAR-1999 (first entry)
DE Enterococcus faecalis genome contig SEQ ID NO:157.
KW Enterococcus faecalis; contig; detection; Enterococcal infection;
KW vaccine; attenuation; computer readable medium; ds.
OS Enterococcus faecalis.
PN WO9850555-A2.
PD 12-NOV-1998.
PF 04-MAY-1998; U08985.
PR 14-NOV-1997; US-066009.
PR 06-MAY-1997; US-044031.
PR 16-MAY-1997; US-046655.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Dillon PJ, Kunsch CA;
DR WPI; 99-045171/04.
PT New isolated Enterococcus faecalis polynucleotides and polypeptides
PT - used to develop products for the detection of Enterococcus and for
PT use in vaccines for prevention or attenuation of Enterococcus
PT infection.
PS Claim 1; Page 888-889; 2084pp; English.
CC A computer readable medium has been developed which has recorded on it
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC X12938 to X13919 represent these nucleotide sequences which are primary
CC nucleotide sequences, also known as contigs. The computer-based system
CC can identify fragments of the Enterococcus faecalis genome with
CC commercial importance. The products can be used to detect the presence
CC of Enterococcus faecalis in samples. They can also be used for
CC diagnosing Enterococcal infection in an animal and monitoring
CC progression of disease, and for identifying agents which can be used to
CC modulate the growth or pathogenicity of Enterococcus faecalis, or
CC another related organism, in vivo or in vitro. In particular the
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC can be used in vaccines to prevent or attenuate an Enterococcal
CC infection.
SQ Sequence 1014 BP; 276 A; 243 C; 182 G; 305 T;

alignment_scores:
Quality: 60.50 Length: 55
Ratio: 1.833 Gaps: 1
Percent Similarity: 60.000 Percent Identity: 32.727

alignment_block:
US-09-068-804-8 x X13094/rev ..

Align seg 1/1 to reverse of: X13094 from: 1 to: 1014

```
      3 ThrSerValArgThrGlnProProValIleMetProGlyMetGlnThrGl 19
        |||:::|||   :::::::::   ||||||      ::::::   :::
    495 ACACCAGTTAGCGTGAATCAAGGCGTGATTGGCGTAGCCTTGGCACCAAC 446

     19 uIleLysThrGlnAlaThrAsnLeuAlaAlaAsnLeuSerAlaValArgG 36
        |||      |||:::::::::||||||   :::|||
    445 AATTGAAGGCCAAACGAAGGACCTGGCAATGGAACTT..... 409

     36 luSerAlaThrAlaThrLeuSerGlyGluIleLysGlyProGlnLeuGlu 52
           ||||||      |||      |||||||:::   :::||||||||||
```

```
     408 ..CTTGCTACTTTAGGGTTAGTCAAAGAAATTCGTGAAGCGCAACTGGAA 361

      53 AspPheProAlaLeu 57
            |||   ::::::
     360 ACTTTTGGTACTGTG 346

seq_name: N_Geneseq_36:Q22689

seq_documentation_block:
ID   Q22689 standard; DNA; 1400 BP.
AC   Q22689;
DT   24-JUL-1992  (first entry)
DE   Sequence encoding acylamino acid racemase.
KW   Enzyme; optically active amino acid; ss.
OS   Amycolatopsis sp. TS-1-60.
FH   Key             Location/Qualifiers
FT   cds             62. .1192
FT                   /*tag=  a
FT                   /codon= (seq:"gtg", aa:Met)
PN   EP-474965-A.
PD   18-MAR-1992.
PF   04-MAY-1991; 107248.
PR   14-SEP-1990; JP-245257.
PA   (TAKE ) TAKEDA CHEMICAL IND KK.
PI   Tokuyama M, Hatano K, Nakahama K, Takahashi T;
DR   WPI; 92-089740/12.
DR   P-PSDB; R21838.
PT   DNA encoding acylamino acid racemase - allows high efficiency
PT   industrial prodn. of enzyme
PS   Claim 6; Fig 1; 22pp; English.
CC   The inventors claim the gene for acylamino acid racemase, esp. bases
CC   62-1189 or 1-1400.  Acylamino acid racemase is a useful enzyme in
CC   the prodn. of optically active amino acids.
SQ   Sequence   1400 BP;    218 A;    465 C;    503 G;    214 T;


alignment_scores:
            Quality:   59.00            Length:     41
              Ratio:  1.788              Gaps:      1
 Percent Similarity:  80.488   Percent Identity: 31.707

alignment_block:
 US-09-068-804-8 x Q22689   ..

 Align seg 1/1  to: Q22689  from: 1  to: 1400

      13 MetProGlyMetGlnThrGluIleLysThrGlnAlaThrAsnLeuAlaAl 29
         |||:::   :::::::::|||:::::::::::::::   ::::::|||||
     374 ATGGCCGTGCTCGACGCCGAACTCCGCGCGCACGAGAGGTCGTTCGCCGC 423

      29 aAsnLeuSerAlaValArgGluSerAlaThrAlaThrLeuSerGlyGluI 46
         |:::|||::::::||||||:::|||::::::      :::|||      |
     424 CGAACTCGGATCGGTGCGCGATTCTGTGCCGTGCGGCGTTTCGGTCGGGA 473

      46 leLysGly......ProGlnLeu 51
         ||:::::::      |||||||||
     474 TCATGGACACCATCCCGCAACTG 496

seq_name: N_Geneseq_36:V77848

seq_documentation_block:
ID   V77848 standard; DNA; 424 BP.
AC   V77848;
DT   16-MAR-1999  (first entry)
DE   Staphylococcus aureus contig SEQ ID #3537.
KW   Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW   cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW   skin infection; surgical wound infection; scalded skin syndrome;
KW   toxic shock syndrome; ds.
OS   Staphylococcus aureus.
PN   EP-786519-A2.
PD   30-JUL-1997.
PF   07-JAN-1997; 100117.
PR   05-JAN-1996; US-009861.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI   Rosen CA;
DR   WPI; 97-374922/35.
PT   Polynucleotide(s) and proteins derived from Staphylococcus aureus  -
PT   stored on computer readable medium and used in the production of
PT   anti-S.aureus vaccines
PS   Claim 1; Page 2573; 3271pp; English.
CC   This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC   of the invention. The DNA sequences are recorded on a computer readable
CC   medium, preferably selected from a floppy or hard disk, random access
CC   memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC   the S.aureus DNA sequences allows putative functions to be assigned so
CC   that protein-encoding or regulatory regions of commercial, therapeutic or
CC   industrial importance can be obtained. Specifically, sequences which are
CC   likely to encode antigens have been identified and these polypeptides can
CC   be used in a vaccine composition against S.aureus infection. The
CC   polypeptides can also be used in a kit for the immunodetection of
CC   S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC   including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC   skin and surgical wound infections, scalded skin syndrome, toxic shock
CC   syndrome, etc. Organisms transformed with the DNA sequences can be used
CC   for recombinant production of the polypeptides. The new DNA sequences
CC   (and their fragments) are useful as primers or probes for isolating
CC   homologues of any of the S.aureus DNA sequences contained on the
CC   computer readable medium.
SQ   Sequence   424 BP;    136 A;    77 C;    97 G;    111 T;


alignment_scores:
            Quality:   54.00            Length:     56
              Ratio:  1.588              Gaps:      1
 Percent Similarity:  60.714   Percent Identity: 30.357

alignment_block:
 US-09-068-804-8 x V77848   ..

 Align seg 1/1  to: V77848  from: 1  to: 424

       8 GlnProProValIleMetProGlyMetGlnThrGluIleLysThrGlnAl 24
         ||||||      :::   ::::::|||      ||||||  ||||||::
      10 CAGCCAGAAGATTTATTCAGTGCGATGATTCGTGAAATTGAAACGCAAGA 59

      24 aThrAsnLeuAlaAlaAsnLeuSerAlaValArgGluSerAlaThrAlaT 41
         :   :::::::::         :::|||:::|||      :::||||:::|
      60 TTTCGATATCGAACACCTGGCGACGGCAATTCGTAAMGTTGAAACATCAA 109

      41 hrLeuSerGlyGluIleLysGlyProGlnLeuGluAspPheProAlaLeu 57
         |||||:::  |||                ::::||||||  :::|||
     110 CATTAGGTGAAGAAAGTGAA.....AATGACTTTATCGGTCTG 147

      58 IleLysGlnAlaSerLeu 63
         :::   :::   :::|||
     148 TTCAGCGATATGGATTTG 165

seq_name: N_Geneseq_36:V77883

seq_documentation_block:
ID   V77883 standard; DNA; 400 BP.
AC   V77883;
DT   16-MAR-1999  (first entry)
DE   Staphylococcus aureus contig SEQ ID #3572.
KW   Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW   cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW   skin infection; surgical wound infection; scalded skin syndrome;
KW   toxic shock syndrome; ds.
OS   Staphylococcus aureus.
PN   EP-786519-A2.
PD   30-JUL-1997.
PF   07-JAN-1997; 100117.
PR   05-JAN-1996; US-009861.
PA   (HUMA-) HUMAN GENOME SCI INC.
```

```
PI   Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI   Rosen CA;
DR   WPI; 97-374922/35.
PT   Polynucleotide(s) and proteins derived from Staphylococcus aureus -
PT   stored on computer readable medium and used in the production of
PT   anti-S.aureus vaccines
PS   Claim 1; Page 2592; 3271pp; English.
CC   This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC   of the invention. The DNA sequences are recorded on a computer readable
CC   medium, preferably selected from a floppy or hard disk, random access
CC   memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC   the S.aureus DNA sequences allows putative functions to be assigned so
CC   that protein-encoding or regulatory regions of commercial, therapeutic or
CC   industrial importance can be obtained. Specifically, sequences which are
CC   likely to encode antigens have been identified and these polypeptides can
CC   be used in a vaccine composition against S.aureus infection. The
CC   polypeptides can also be used in a kit for the immunodetection of
CC   S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC   including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC   skin and surgical wound infections, scalded skin syndrome, toxic shock
CC   syndrome, etc. Organisms transformed with the DNA sequences can be used
CC   for recombinant production of the polypeptides. The new DNA sequences
CC   (and their fragments) are useful as primers or probes for isolating
CC   homologues of any of the S.aureus DNA sequences contained on the
CC   computer readable medium.
SQ   Sequence   400 BP;    131 A;    66 C;    93 G;    110 T;


alignment_scores:
            Quality:   52.00              Length:     56
              Ratio:  1.486                Gaps:      1
 Percent Similarity:  62.500   Percent Identity: 28.571

alignment_block:
 US-09-068-804-8 x V77883    ..

 Align seg 1/1  to: V77883  from: 1  to: 400

       8 GlnProProValIleMetProGlyMetGlnThrGluIleLysThrGlnAl 24
         :::|||:::   :::   ::::::|||      ||||||   ||||||::
      93 GAGCCACAAGATTTATTCAGTGCGATGATTCGTGAAATTGAAACGCAAGA 142

      24 aThrAsnLeuAlaAlaAsnLeuSerAlaValArgGluSerAlaThrAlaT 41
         :   :::::::::          :::|||:::|||      :::|||:::|
     143 TTTCGATATAGAACATCTAGCGACGGCGATTCGCAAAGTTGAAACATCTA 192

      41 hrLeuSerGlyGluIleLysGlyProGlnLeuGluAspPheProAlaLeu 57
         |||||:::   |||                  :::||||||   :::|||
     193 CATTAGGTGAAGAAAGTGAA.............AATGACTTTATCGGGCTG 230

      58 IleLysGlnAlaSerLeu 63
         :::   :::   :::|||
     231 TTCAGCGATATGGATTTG 248

seq_name: N_Geneseq_36:N92320

seq_documentation_block:
ID   N92320 standard; DNA; 219 BP.
AC   N92320;
DT   1-JUL-1990  (first entry)
DE   Murine autonomously replicating sequence 9 (MuARS-9)
KW   Murine autonomously replicating sequence 9; MARS 9; MuARS-9;
KW   consensus sequence; mammalian expression vector pBR327;
KW   plasmid pTK (DSM 4203P); mouse LMTK- cells.
OS   Mouse.
PN   EP-306848-A.
PD   15-MAR-1989.
PF   02-SEP-1988; 114317.
PR   09-SEP-1987; DE-730246.
PA   (BOEF) Boehringer Mannheim GMBH.
PI   Grummt F, Weidle U;
DR   WPI; 89-078357/11.
PT   Expression vectors for mammalian cells -
PT   with two consensus sequences and inefficient selection system
PS   Table 2; page 7; 12pp; German.
CC   The expression vectors of the patent contain a consensus sequence (CS).
CC   The CS or their homologues are present in known MuARS. The
CC   inefficient selection system comprises the tk gene and a truncated tk
CC   promoter. The vectors also contain a gene coding for a protein and the
CC   necessary promoter and terminator sequences for the expression of this
CC   gene. The expression vectors provide rapid amplification,
CC   allow selection of best clones in a short time, are of universal
CC   utility in mammalian cells, and are free of viral origin sequences.
CC   In the example, a 2487 bp EcoRI fragment of herpes simplex virus I
CC   (HSVI) DNA contg. the complete tk gene and 80 bp of tk promoter was
CC   inserted into pBR327 to produce plasmid pTK (DSM 4203P). This was
CC   ligated with murine autonomously replicating sequence (MuARS) 1. The
CC   recombinant plasmid was cloned in E. coli HB101 and used to transfect
CC   mouse LMTK- cells. After culturing the cells for 60 days, the replication
CC   rate was 530+/-270 copies/cell.
SQ   Sequence   219 BP;    53 A;    45 C;    49 G;    72 T;


alignment_scores:
            Quality:   48.00              Length:     28
              Ratio:  2.286                Gaps:      0
 Percent Similarity:  75.000   Percent Identity: 35.714

alignment_block:
 US-09-068-804-8 x N92320    ..

 Align seg 1/1  to: N92320  from: 1  to: 219

       2 ValThrSerValArgThrGlnProProValIleMetProGlyMetGlnTh 18
         :::|||:::|||||||||      ::::::::::|||:::   :::::::||
     118 ATTACTGGTGTAAGGACTCTGGTTTCCATGTTTATGTCTACCTTGAGAAC 167

      18 rGluIleLysThrGlnAlaThrAsnLeuAlaAla 29
         ||||:::      |||::::::   |||   |||
     168 TGAGTTAGATCAGCAGGACCCACCTCTCTGTGCC 201

seq_name: N_Geneseq_36:X00509

seq_documentation_block:
ID   X00509 standard; DNA; 704 BP.
AC   X00509;
DT   22-APR-1999  (first entry)
DE   Rat EGR-1 gene promoter.
KW   Replication-defective recombinant virus; eye surgery;
KW   lens epithelial cell; secondary cataract prevention; rat;
KW   early growth response-1; EGR-1 gene promoter; ss.
OS   Rattus sp.
PN   WO9855641-A2.
PD   10-DEC-1998.
PF   03-JUN-1998; IB0862.
PR   03-JUN-1997; US-867902.
PA   (MUNI/) MUNIER F.
PA   (SAHL/) SAHLI R.
PA   (SHAW/) SHAW P H.
PA   (SICK/) SICKENBERG M.
PI   Munier F, Sahli R, Shaw PH, Sickenberg M;
DR   WPI; 99-070276/06.
PT   Replication defective recombinant virus capable of infecting the eye
PT   - used to prevent formation of secondary cataracts in eye surgery
PS   Example 1; Fig 4; 65pp; English.
CC   This sequence represents the rat early growth response-1 (EGR-1) gene
CC   promoter. This sequence can be used in the replication-defective
CC   recombinant virus of the invention, which is capable of infecting an eye,
CC   and contains an open reading frame (ORF) encoding a protein that when
CC   expressed in lens epithelial cells of an eye, suppresses, at the level of
CC   germinative epithelium of the lens of the eye, cellular proliferation
CC   which is stimulated by eye surgery and which would otherwise result in
CC   secondary cataract formation in the eye, the ORF, to be expressed, is
CC   under the control of a promoter sequence which is active exclusively in
CC   human lens epithelial cells. The replication defective recombinant virus
CC   can be used in a method for the treatment of an eye of a patient
```

```
CC   undergoing eye surgery, and thereby prevent the formation of secondary
CC   cataracts.
SQ   Sequence   704 BP;    112 A;    241 C;    237 G;    114 T;


alignment_scores:
            Quality:   52.00              Length:     45
              Ratio:  1.625                Gaps:      1
 Percent Similarity:  71.111   Percent Identity: 33.333

alignment_block:
 US-09-068-804-8 x X00509   ..

 Align seg 1/1  to: X00509  from: 1  to: 704

       6 ArgThrGlnProProValIleMetProGlyMetGlnThrGluIleLysTh 22
         |||      ||||||         |||:::      ::::::      ::
     568 CGCGATTCGCCGCCGCCG......CCAGCTTCCGCCGCCGCAAGATCGGC 611

      22 rGlnAlaThrAsnLeuAlaAlaAsnLeuSerAlaValArgGluSerAlaT 39
         ::::|||:::::   ||||||:::|||   :::   |||:::::|||| 
     612 CCCTGCCCCAGCCTCCGCGGCAGCCCTGCGTCCACCACGGGCCGCGGCCA 661

      39 hrAlaThrLeuSerGlyGluIleLysGlyProGln 50
         |||||:::|||:::::::::::::   :::|||||| 
     662 CCGCCAGCCTGGGGGCCCACCTACACTCCCCGCAG 696

seq_name: N_Geneseq_36:Q29170

seq_documentation_block:
ID   Q29170 standard; DNA; 117 BP.
AC   Q29170;
DT   09-MAR-1993  (first entry)
DE   HLA-Bw 52 exon 5.
KW   Human leukocyte antigen; transgenic; germ cells; somatic cells;
KW   expression; ss.
PN   J04091731-A.
PD   25-MAR-1992.
PF   03-AUG-1990; 207329.
PR   03-AUG-1990; JP-207329.
PA   (OLYU ) OLYMPUS OPTICAL CO.
DR   WPI; 92-342893/42.
PT   Transgenic non-human mammalian HLA-Bw 52 gene - useful for
PT   analysis of expression of gene structure, and prodn. of
PT   mouse model of human disease
PS   Disclosure; Fig 1; 8pp; Japanese.
CC   The sequence shows exon 5 of the human leukocyte antigen-Bw 52 gene.
CC   The complete gene may be introduced into non-human mammals, pref.
CC   rat or mouse, or their ancestors at the primary developmental
CC   biological step via transplantation into the zygote or embryo to
CC   generate transgenic non-human mammals incorporating the HLA-Bw 52
CC   gene in both their germ cells and somatic cells.  Transgenic non-human
CC   mammals contg. HLA-Bw 52 are useful for the analysis of expression of
CC   the gene, its structure, and prodn. of mouse models of human disease.
CC   See also Q29166-72.
SQ   Sequence   117 BP;    19 A;    30 C;    37 G;    31 T;


alignment_scores:
            Quality:   45.00              Length:     16
              Ratio:  3.462                Gaps:      0
 Percent Similarity:  81.250   Percent Identity: 50.000

alignment_block:
 US-09-068-804-8 x Q29170/rev   ..

 Align seg 1/1  to reverse of: Q29170  from: 1  to: 117

       1 MetValThrSerValArgThrGlnProProValIleMetProGlyMet 16
         |||:::|||::::::||||||      |||:::::::||||||   |||
      73 ATGACCACAACTGCTAGGACAGCCAGGCCAGCAACAATGCCCACGATG 26

seq_name: N_Geneseq_36:V89595

seq_documentation_block:
ID   V89595 standard; cDNA; 600 BP.
AC   V89595;
DT   15-FEB-1999  (first entry)
DE   EST clone CP147.
KW   Human; secreted protein; expressed sequence tag; EST; haematopoiesis;
KW   tissue growth; activin; inhibin; chemotaxis; chemokinesis; haemostatic;
KW   receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;
KW   gene therapy; ss.
OS   Homo sapiens.
PN   WO9845436-A2.
PD   15-OCT-1998.
PF   10-APR-1998; U06955.
PR   10-APR-1997; US-838821.
PA   (GEMY ) GENETICS INST INC.
PI   Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
PI   Racie LA, Spaulding V, Treacy M;
DR   WPI; 99-070077/06.
PT   New polynucleotides encoding human secreted proteins - derived from
PT   e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT   ovary, pituitary, retina and colon cDNA libraries.
PS   Claim 1; Page 268; 618pp; English.
CC   The present sequence represents a human expressed sequence tag (EST).
CC   The polynucleotide, which is a secreted EST, and the encoded protein
CC   are predicted to have useful biological activities which would make
CC   them suitable for treating, preventing or ameliorating medical
CC   conditions in humans and animals, although no supporting data is
CC   given. Suggested activities include nutritional activity, immune
CC   stimulating or suppressing activity, haematopoiesis regulating
CC   activity, tissue growth activity, activin/inhibin activity,
CC   chemotactic/chemokinetic activity, haemostatic and thrombolytic
CC   activity, receptor/ligand activity, anti-inflammatory activity,
CC   cadherin/tumour invasion suppressor activity, tumour inhibition
CC   activity. The polynucleotide may also be useful for gene therapy.
SQ   Sequence   600 BP;    153 A;    153 C;    207 G;    87 T;


alignment_scores:
            Quality:   51.00              Length:     39
              Ratio:  1.759                Gaps:      0
 Percent Similarity:  74.359   Percent Identity: 30.769

alignment_block:
 US-09-068-804-8 x V89595   ..

 Align seg 1/1  to: V89595  from: 1  to: 600

      12 IleMetProGlyMetGlnThrGluIleLysThrGlnAlaThrAsnLeuAl 28
         :::|||   :::     ||||||:::|||:::      :::::::|||:: 
     283 CTGATGGAGGACACTATGACGGAAGTAAAGGCTTACAAAAAGGAGCTGGA 332

      28 aAlaAsnLeuSerAlaValArgGluSerAlaThrAlaThrLeuSerGlyG 45
         :::::::|||::::::|||   |||   :::   |||   |||:::   | 
     333 GGAACAGCTGGGTCCAGTGGCGGAGGAGACACGGGCCAGGCTGGGCAAAG 382

      45 luIleLysGlyProGln 50
         ||:::::::::::::|||
     383 AGGTGCAGGCGGCACAG 399

seq_name: N_Geneseq_36:Q12072

seq_documentation_block:
ID   Q12072 standard; DNA; 2100 BP.
AC   Q12072;
DT   29-AUG-1991  (first entry)
DE   T-PA with -ve charged finger and/or kringle domain (2).
KW   T-PA; tissue plasminogen activator; finger; kringle domain;
KW   thrombosis; lysis; ss.
FH   Key             Location/Qualifiers
FT   cds             24. .1709
FT                   /*tag=  a
FT                   /product= t-PA variant
```

```
FT   misc_rna        120. .131
FT                   /*tag= b
FT                   /label= R1
FT   misc_rna        132. .146
FT                   /*tag= c
FT                   /label= A2-6
FT                   /note= "Tyr2-Cys6 of natural t-PA"
FT   misc_rna        147. .218
FT                   /*tag= d
FT                   /label= R2
FT   misc_rna        219. .497
FT                   /*tag= e
FT                   /label= A31-123
FT                   /note= "VAl31-Gln123 of natural t-PA"
FT   misc_rna        498. .533
FT                   /*tag= f
FT                   /label= R3
FT   misc_rna        534. .1709
FT                   /*tag= g
FT                   /label= A136-527
FT                   /note= "Leu136-Pro527"
PN   J03061483-A.
PD   18-MAR-1991.
PF   26-JUN-1989; 163599.
PR   27-JUN-1988; GB-015246.
PA   (FUJI ) FUJISAWA PHARM KK.
DR   WPI; 91-180545/25.
DR   P-PSDB; R123432.
PT   New tissue plasminogen activator with negatively charged finger
PT   and/or kringle domain - is highly active and is useful as thrombosis-
PT   lysing agent
PS   Disclosure; Fig 28; 29pp; Japanese.
CC   This sequence is an example of a generic formula.  The activator
CC   is highly active in converting plasminogen to plasmin.  It is
CC   useful as a thrombosis-lysing agent and is used in a pharmaceutical
CC   compsn. to treat diseases, such as cardiac infarction, cerebral
CC   haemorrhage, heart stroke and pulmonary embolism.
CC   See Q12071-2, R13150-53 and R12366-67 and also J03061482-84.
SQ   Sequence   2100 BP;    497 A;    595 C;    598 G;    410 T;


alignment_scores:
            Quality:   55.50              Length:     58
              Ratio:  1.632                Gaps:      3
 Percent Similarity:  58.621   Percent Identity: 36.207

alignment_block:
 US-09-068-804-8 x Q12072   ..

 Align seg 1/1  to: Q12072  from: 1  to: 2100

       3 ThrSerValArgThrGlnProProVal..............IleMetPr 14
         |||:::   |||:::|||:::|||                 :::   ::
     505 ACAGCGGGGAGGAGCCAGACGCCCTCGAGCTGGGCCTGGGGAACCACAAC 554

      14 oGlyMetGlnThrGluIleLysThrGln.........AlaThrAsnLeu. 27
         :      :::|||:::|||   ||||||         ||||||:::|||
     555 TACTGCAGAAACCCAGATCGAGACTCAAAGCCCTGGTGCTACGTCTTTAA 604

      28 ........................AlaAlaAsnLeuSerAlaValArg 35
                                 ||||||   |||:::|||:::|||
     605 GGCGGGGAAGTACAGCTCAGAGTTCTGCAGCACCCCTGCCTGCTCTGAGG 654

      36 GluSerAlaThrAlaThrLeuSer 43
         |||::::::|||||||||||:::
     655 GAAACAGTGACTGCTACTTTGGGA 678

seq_name: N_Geneseq_36:T98639

seq_documentation_block:
ID   T98639 standard; DNA; 1568 BP.
AC   T98639;
DT   09-NOV-1998  (first entry)
DE   DNA encoding a S. pneumoniae recombination protein.
KW   Streptococcus pneumoniae protein; genetic immunisation; antagonist;
KW   immunological response; inoculation; antibody production; inhibitor;
KW   T cell immune response; antimicrobial compound; bacterial adhesion;
KW   extracellular matrix protein; protein-mediated cell invasion; wound;
KW   pathogenesis; ss.
OS   Streptococcus pneumoniae.
FH   Key             Location/Qualifiers
FT   CDS             898. .1527
FT                   /*tag= a
PN   WO9743303-A1.
PD   20-NOV-1997.
PF   14-MAY-1997; U07950.
PR   14-MAY-1996; US-017670.
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PA   (SMIK ) SMITHKLINE BEECHAM PLC.
PI   Black MT, Hodgson JE, Knowles DJC, Nicholas RO,
PI   Stodola RK;
DR   WPI; 98-008793/01.
DR   P-PSDB; W38589.
PT   Novel Streptococcus pneumoniae proteins and related DNA - useful for
PT   diagnosing anti-microbial agents for treatment of bacterial
PT   infections
PS   Claim 4; Page 164; 483pp; English.
CC   This sequence encodes a Streptococcus pneumoniae protein that (based on
CC   homology with a Bacillus subtilis protein) is a recombination
CC   protein, and represents a DNA sequence of the invention.
CC   The DNA sequences were isolated from Streptococcus pneumoniae strain
CC   0100993 (NCIMB 40794). The Streptococcus pneumoniae proteins of the
CC   invention can be used to identify compounds which interact with and
CC   inhibit or activate the activity of the proteins. Antagonists can be
CC   used to treat diseases caused by S. pneumoniae proteins, through genetic
CC   immunisation. They can also be used to induce an immunological response
CC   in a mammal by inoculation with the S. pneumoniae proteins or delivery
CC   of the encoding nucleic acids in a vector adequate to produce antibody
CC   and/or T cell immune responses to protect the animal from disease. The
CC   proteins can also be used to identify antimicrobial compounds which are
CC   capable of inhibiting their bioactivity. In particular the proteins of
CC   the invention can be used to prevent adhesion of bacteria to mammalian
CC   extracellular matrix proteins on in-dwelling devices or in wounds, to
CC   block protein-mediated mammalian cell invasion, and to block the normal
CC   progression of pathogenesis in infections initiated other than by the
CC   implantation of in-dwelling devices or other surgical techniques.
SQ   Sequence   1568 BP;    462 A;    269 C;    385 G;    452 T;


alignment_scores:
            Quality:   54.00              Length:     50
              Ratio:  1.421                Gaps:      3
 Percent Similarity:  76.000   Percent Identity: 32.000

alignment_block:
 US-09-068-804-8 x T98639   ..

 Align seg 1/1  to: T98639  from: 1  to: 1568

      10 ProValIleMetProGly......MetGlnThrGluIleLys........ 21
         ::::::|||   |||:::      |||::::::|||:::|||
     874 GCAATAATCTTTCCGTCTGAGGACATGGAAGCAGAGCTTAAGAAGTTGGA 923

      22 .ThrGlnAlaThrAsnLeuAlaAlaAsnLeuSerAlaValArgGluSerA 38
          ::::::   :::||||||||||::::::|||::::::::::|||::::::
     924 AGTCAATCTTGTCAATTTGGCAGGTCAACTTGCTTCTGCTCGTCATAATT 973

      38 laThrAlaThrLeuSerGlyGluIleLysGlyProGlnLeuGluAspPhe 54
           :::      |||   :::|||||||||   :::::::|||:::|||:::
     974 TGGCTCAGCAACTCGAAGCTGAGATTAAA...CAAGAACTGCAAGATCTT 1020

seq_name: N_Geneseq_36:T22145

seq_documentation_block:
ID   T22145 standard; cDNA to mRNA; 379 BP.
AC   T22145;
```

```
DT   02-AUG-1996  (first entry)
DE   Human gene signature HUMGS03715.
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
OS   Homo sapiens.
PN   WO9514772-A1.
PD   01-JUN-1995.
PF   11-NOV-1994; J01916.
PR   12-NOV-1993; JP-355504.
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
PI   Matsubara K,  Okubo K;
DR   WPI; 95-206931/27.
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
PS   Claim 1; Page 1054; 2245pp; Japanese.
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
SQ   Sequence   379 BP;    92 A;    84 C;    105 G;    97 T;


alignment_scores:
            Quality:   48.00              Length:     27
              Ratio:  2.400                Gaps:      0
 Percent Similarity:  74.074   Percent Identity: 44.444

alignment_block:
 US-09-068-804-8 x T22145/rev   ..

 Align seg 1/1  to reverse of: T22145  from: 1  to: 379

       7 ThrGlnProProValIleMetProGlyMetGlnThrGluIleLysThrGl 23
         |||||||||||||||      |||      |||::::::   |||:::||
     135 ACTCAGCCTCCAGTCCCCCGCCCCATAGGTCAGGCTGACTCAAAGCCTCA 86

      23 nAlaThrAsnLeuAlaAlaAsnLeuSerAla 33
         |:::::::|||      ::::::||||||:::
      85 AACCTCTAATCACATGGTCAGTCTGTCTGGT 55

seq_name: N_Geneseq_36:V77996

seq_documentation_block:
ID   V77996 standard; DNA; 400 BP.
AC   V77996;
DT   16-MAR-1999  (first entry)
DE   Staphylococcus aureus contig SEQ ID #3685.
KW   Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW   cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW   skin infection; surgical wound infection; scalded skin syndrome;
KW   toxic shock syndrome; ds.
OS   Staphylococcus aureus.
PN   EP-786519-A2.
PD   30-JUL-1997.
PF   07-JAN-1997; 100117.
PR   05-JAN-1996; US-009861.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI   Rosen CA;
DR   WPI; 97-374922/35.
PT   Polynucleotide(s) and proteins derived from Staphylococcus aureus  -
PT   stored on computer readable medium and used in the production of
PT   anti-S.aureus vaccines
PS   Claim 1; Page 2643; 3271pp; English.
CC   This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC   of the invention. The DNA sequences are recorded on a computer readable
CC   medium, preferably selected from a floppy or hard disk, random access
CC   memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC   the S.aureus DNA sequences allows putative functions to be assigned so
CC   that protein-encoding or regulatory regions of commercial, therapeutic or
CC   industrial importance can be obtained. Specifically, sequences which are
CC   likely to encode antigens have been identified and these polypeptides can
CC   be used in a vaccine composition against S.aureus infection. The
CC   polypeptides can also be used in a kit for the immunodetection of
CC   S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC   including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC   skin and surgical wound infections, scalded skin syndrome, toxic shock
CC   syndrome, etc. Organisms transformed with the DNA sequences can be used
CC   for recombinant production of the polypeptides. The new DNA sequences
CC   (and their fragments) are useful as primers or probes for isolating
CC   homologues of any of the S.aureus DNA sequences contained on the
CC   computer readable medium.
SQ   Sequence   400 BP;    129 A;    61 C;    96 G;    111 T;


alignment_scores:
            Quality:   48.00              Length:     38
              Ratio:  1.846                Gaps:      0
 Percent Similarity:  68.421   Percent Identity: 31.579

alignment_block:
 US-09-068-804-8 x V77996   ..

 Align seg 1/1  to: V77996  from: 1  to: 400

       8 GlnProProValIleMetProGlyMetGlnThrGluIleLysThrGlnAl 24
         :::|||:::   :::   ::::::|||      ||||||   ||||||::
     226 GAGCCACAAGATTTATTCAGTGCGATGATTCGTGAAATTGAAACGCAAGA 275

      24 aThrAsnLeuAlaAlaAsnLeuSerAlaValArgGluSerAlaThrAlaT 41
         :   :::::::::         :::|||:::|||      :::|||:::|
     276 TTTCGATATAGAGCATCTAGCGACGGCGATTCGCAAAGTTGAAACATCTA 325

      41 hrLeuSerGlyGlu 45
         |||||:::   |||
     326 CATTAGGTGANGAA 339

seq_name: N_Geneseq_36:T05399

seq_documentation_block:
ID   T05399 standard; DNA; 540 BP.
AC   T05399;
DT   04-JUN-1996  (first entry)
DE   Mycosphaerella musicola internal transcribed spacer sequence.
KW   Plant pathogen; fungus; Septoria nodorum; Septoria tritici; Fusarium;
KW   Pseudocercosporella herpotrichoides; Mycosphaerella fijiensis; PCR;
KW   Mycosphaerella musicola; amplification; primer; ribosomal RNA gene;
KW   internal transcribed region; strain; capture; colourimetric assay;
KW   isolate; development; population; random amplified polymorphic DNA; ss.
OS   Mycosphaerella musicola.
PN   WO9529260-A2.
PD   02-NOV-1995.
PF   19-APR-1995; U04712.
PR   25-APR-1994; US-233608.
PA   (CIBA ) CIBA GEIGY AG.
PI   Beck JJ,  Ligon JM;
DR   WPI; 95-383005/49.
PT   DNA encoding intervening transcribed sequence  - used for detection
PT   of plant fungal pathogens
PS   Claim 1; Page 30; 65pp; English.
CC   A novel method for the detection of plant pathogenic strains of fungi
```

```
CC   e.g. Septoria nodorum, S.tritici, Pseudocercosporella herpotrichoides,
CC   Mycosphaerella fijiensis, M.musicola or Fusarium spp, involves the PCR
CC   amplification of sequences found in the internal transcribed region
CC   (ITS) of the 18S, 5.8S and 28S ribosomal RNA genes by the primers
CC   Q94359-93 and T05357-72.  These primers are derived from the ITS
CC   sequences of these fungi (T05394-T05404 and Q94398) and are strain
CC   specific.  The amplification products of the reactions using these
CC   primers can be used with the capture primers T05378-93 in colourimetric
CC   assays.  The primers and ITS DNAs can be used for the detection of
CC   specific fungal pathogen isolates and in monitoring disease development
CC   in plant populations.
SQ   Sequence   540 BP;    117 A;    154 C;    146 G;    123 T;


alignment_scores:
            Quality:   49.00              Length:     45
              Ratio:  1.531                Gaps:      1
 Percent Similarity:  71.111   Percent Identity: 26.667

alignment_block:
 US-09-068-804-8 x T05399/rev   ..

 Align seg 1/1  to reverse of: T05399  from: 1  to: 540

       6 ArgThrGlnProProValIleMetProGlyMetGlnThrGluIleLysTh 22
         |||:::::|||||:::::   :::   :::::::::      :::
     467 CGGCCGCGGCCACCCGCGCTCCGGAGCGAACTGAAAAATCCACAACGCTT 418

      22 rGlnAlaThrAsnLeuAlaAlaAsnLeuSerAlaValArgGluSerAlaT 39
          ::::::   :::|||||:::   |||   ::::::|||   :::|||:
     417 AGAGACGGACGGCTTAGCCGGGAGACTTTGGGGCGCGCGGAGCACCGCGG 368

      39 hrAlaThr...........LeuSerGlyGluIle 46
         ::::::::           ||||||||||||:::
     367 CGCCCAATACCAAGCTAGGCTTGAGTGGTGAAATG 333
```

OM of: US-09-068-804-8 to: GenEmbl:* out_format : pfs

Date: Oct 5, 1999 8:00 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09068804/runat_05101999_145134_10315/app_query.fasta.1
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=pam150 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=escore -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfs -NORM=stat -USER=US09068804 -NCPU=6 -ICPU=3 -WAIT
-THREADS=1

Search information block:
Query: US-09-068-804-8
Query length: 64
Database: GenEmbl:*
Database sequences: 679419
Database length: 1590154680
Search time (sec): 1424.920000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ..
gb_ba2:STU30491	+	277.00	663.44	5.2e-29	3441	! U30491 Salmonella typhimurium
gb_ba1:STSIPGNS	+	271.00	644.40	5.9e-28	5393	! X82670 S.typhi spaT, sipB, sip
gb_ba1:STU40013	+	235.00	558.50	3.6e-23	3500	! U40013 Salmonella typhimurium
gb_pl1:YLSEC62	+	62.00	136.19	12.07	1256	! X99537 Y.lipolytica SEC62 gene
gb_pl1:WHTGLUTB	+	64.00	133.28	17.53	2999	! M22209 Tricticum aestivum, hig
gb_ba1:PAARCC	-	61.00	133.06	18.01	1346	! X14693 Pseudomonas aeruginosa
gb_ov:EMU87434	-	56.50	132.01	20.63	439	! U87434 Eleginops maclovinus 28S
gb_vi:PPVNIB	+	55.50	131.16	22.99	366	! L42472 Plum pox virus coat prot
gb_pl1:TAGLB11B	+	64.00	130.59	24.75	4034	! X13927 Wheat Glu-B1-1b gene fc
gb_vi:PPVNIBA	+	55.50	130.11	26.31	411	! L42473 Plum pox virus coat prot
gb_in1:LTTRNAGLN	-	56.50	127.82	35.27	696	! X68207 L.tarentolae gene for tR
gb_ba1:AMLAAAR	+	59.00	127.72	35.76	1400	! D30738 Amycolatopsis sp. Aaar
gb_pat:E04131	+	59.00	127.72	35.76	1400	! E04131 DNA sequence encoding A
gb_pat:I22022	+	59.00	127.72	35.76	1400	! I22022 Sequence 1 from patent
gb_pl2:AF000939	-	56.50	127.28	37.82	739	! AF000939 Hordeum vulgare aleurc
gb_pr1:HS31C4F	-	53.00	127.16	38.39	286	! Z65303 H.sapiens CpG island DNA
gb_in1:APLNP	+	58.00	126.47	41.95	1220	! K01223 Aplysia (marine gastrop
gb_ba1:BSU40655	+	59.50	126.15	43.75	1910	! U40655 Burkholderia solanacear
gb_pl2:AF045417	-	54.00	126.13	43.85	422	! AF045417 Phytophthora infestans
gb_vi:PPNLCP	+	58.00	126.11	43.96	1270	! X81074 Plum pox potyvirus gene
gb_pr2:HSY11284	+	56.00	125.42	48.04	791	! Y11284 Homo sapiens AFX1 gene,
gb_ro:RNU80915	-	55.00	124.04	57.29	699	! U80915 Rattus norvegicus EAAT4
gb_ro:AF115768	+	53.00	123.96	57.89	407	! AF115768 Rattus norvegicus defe
gb_ba1:MSNAPS218	-	51.50	123.20	63.81	293	! X98384 Moraxella sp. napA gene,
gb_in1:PTGAPC	-	51.50	122.99	65.59	300	! X94324 Paramecium tetraurelia g
gb_pl1:WHTGLIGP	+	55.00	122.84	66.84	798	! M16060 Wheat (T.aestivum) gamma
gb_vi:PPPVCOAT	+	55.50	122.13	73.21	990	! AJ000340 Plum pox potyvirus mRN
gb_ro:MUSCDA44A	+	56.00	121.81	76.31	1177	! M27129 Mouse CD44 antigen mRNA
gb_ro:MUSCDA44B	+	56.00	121.76	76.76	1183	! M27130 Mouse CD44 antigen mRNA
gb_ro:MUSPGP1A	+	56.00	121.57	78.65	1208	! J05163 Mouse phosphoglycoprote
gb_pl2:AF120267	+	55.00	121.29	81.58	947	! AF120267 Triticum spelta gamma-
gb_pl2:AF144104	+	55.00	121.20	82.49	956	! AF144104 Triticum aestivum gamm
gb_vi:PPLIHCP	+	56.00	121.12	83.37	1270	! X81081 Plum pox potyvirus gene
gb_vi:PPSLCP	+	56.00	121.12	83.37	1270	! X81079 Plum pox potyvirus gene
gb_ba1:AB012859	+	55.50	120.87	86.10	1138	! AB012859 Flexibacter aurantiac
gb_ba1:AB012860	+	55.50	120.87	86.10	1138	! AB012860 Flexibacter aurantiac
gb_pl1:ANRRN5P2	+	51.00	120.82	86.61	332	! X03569 Aspergillus nidulans 5S
gb_ro:MUSPGP1	+	56.00	120.65	88.51	1337	! M30655 Mouse phagocytic glycop
gb_ba1:SPE005575	+	58.00	120.63	88.79	2323	! AJ005575 Streptomyces peucetiu
gb_ov:AF027222	+	54.50	120.43	91.05	907	! AF027222 Cepphus columba adiant
gb_ov:AF027224	+	54.50	120.42	91.16	908	! AF027224 Cepphus columba eureka
gb_ov:AF027250	+	54.50	120.41	91.28	909	! AF027250 Cepphus columba adiant
gb_sts:G30668	-	51.50	120.17	94.09	409	! G30668 human STS SHGC-37722,
gb_ro:MMPGP1M1M	+	56.00	120.12	94.78	1418	! X66081 M.musculus Pgp-1 mRN
gb_vi:PPCOAPROA	+	55.50	119.87	97.83	1270	! X57975 Plum pox potyvirus m

seq_name: gb_ba2:STU30491

seq_documentation_block:
LOCUS STU30491 3441 bp DNA BCT 19-JAN-1996
DEFINITION Salmonella typhimurium SspA (sspA) and SspB (sspB) genes, partial cds, and SspC (sspC) and SspD (sspD) genes, complete cds.
ACCESSION U30491
NID g1117811
VERSION U30491.1 GI:1117811
KEYWORDS .
SOURCE Salmonella typhimurium.
ORGANISM Salmonella typhimurium
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Salmonella.
REFERENCE 1 (bases 1 to 3441)
AUTHORS Hueck,C.J., Hantman,M.J., Bajaj,V., Johnston,C., Lee,C.A. and Miller,S.I.
TITLE Salmonella typhimurium secreted invasion determinants are homologous to Shigella Ipa proteins
JOURNAL Mol. Microbiol. 18 (3), 479-490 (1995)
MEDLINE 96342378
REFERENCE 2 (bases 1 to 3441)
AUTHORS Miller,S.I.
TITLE Direct Submission
JOURNAL Submitted (28-JUN-1995) Samuel I. Miller, Infectious Disease Unit, Massachusetts General Hospital, Harvard Medical School, Fruit Street, Boston, MA 02114, USA
FEATURES Location/Qualifiers
source 1. .3441
/organism="Salmonella typhimurium"
/strain="SL1344"
/db_xref="taxon:602"
/map="59 min"
gene 1. .870
/gene="sspB"
CDS <1. .870
/gene="sspB"
/note="secreted protein"
/codon_start=1
/transl_table=11
/function="invasion determinant"
/product="SspB"
/protein_id="AAC43546.1"
/db_xref="PID:g1117812"
/db_xref="GI:1117812"
/translation="RKAEETNRIMGCIGKVLGALLTIVSVVAAVFTGGASLALAAVGL
AVMVADEIVKAATGVSFIQQALNPIMEHVLKPLMELIGKAITKALEGLGVDKKTAEMA
GSIVGAIVAAIAMVAVIVVVAVVGKGAAAKLGNALSKMMGETIKKLVPNVLKQLAQNG
SKLFTQGMQRITSGLGNVGSKMGLQTNALSKELVGNTLNKVALGMEVTNTAAQSAGGV
AEGVFIKNASEALADFMLARFAMDQIQQWLKQSVEIFGENQKVTAELQKAMSSAVQQN
ADASRFILRQSRA"
gene 898. .2127
/gene="sspC"
CDS 898. .2127
/gene="sspC"
/note="secreted protein"
/codon_start=1
/transl_table=11
/function="invasion determinant"
/evidence=experimental
/product="SspC"
/protein_id="AAC43547.1"
/db_xref="PID:g1117813"
/db_xref="GI:1117813"
/translation="MLISNVGINPAAYLNNHSVENSSQTASQSVSAKDILNSIGISSS
KVSDLGLSPTLSAPAPGVLTQTPGTITSFLKASIQNTDMNQDLNALANNVTTKANEVV
QTQLREQQAEVGKFFDISGMSSSAVALLAAANTLMLTLNQADSKLSGKLSLVSFDAAK
TTASSMMREGMNALSGSISQSALQLGITGVGAKLEYKGLQNERGALKHNAAKIDKLTT
ESHSIKNVLNGQNSVKLGAEGVDSLKSLNMKKTGTDATKNLNDATLKSNAGTSATESL

```
                     GIKDSNKQISPEHQAILSKRLESVESDIRLEQNTMDMTRIDARKMQMTGDLIMKNSVT
                     VGGIAGASGQYAATQERSEQQISQVNNRVASTASDEARESSRKSTSLIQEMLKTMESI
                     NQSKASALAAAIAGNIRA"
     gene            2198. .3229
                     /gene="sspD"
     CDS             2198. .3229
                     /gene="sspD"
                     /note="secreted protein"
                     /codon_start=1
                     /transl_table=11
                     /function="invasion determinant"
                     /evidence=experimental
                     /product="SspD"
                     /protein_id="AAC43548.1"
                     /db_xref="PID:g1117814"
                     /db_xref="GI:1117814"
                     /translation="MLNIQNYSASPHPGIVAERPQTPSASEHVETAVVPSTTEHRGTD
                     IISLSQAATKIHQAQQTLQSTPPISEENNDERTLARQQLTSSLNALAKSGVSLSAEQN
                     ENLRSAFSAPTSALFSASPMAQPRTTISDAEIWDMVSQNISAIGDSYLGVYENVVAVY
                     TDFYQAFSDILSKMGGWLLPGKDGNTVKLDVTSLKNDLNSLVNKYNQINSNTVLFPAQ
                     SGSGVKVATEAEARQWLSELNLPNSCLKSYGSGYVVTVDLTPLQKMVQDIDGLGAPGK
                     DSKLEMDNAKYQAWQSGFKAQEENMKTTLQTLTQKYSNANSLYDNLVKVLSSTISSSL
                     ETAKSFLQG"
     gene            3248. .3441
                     /gene="sspA"
     CDS             3248. .>3441
                     /gene="sspA"
                     /note="secreted protein"
                     /codon_start=1
                     /transl_table=11
                     /function="unknown"
                     /product="SspA"
                     /protein_id="AAC43549.1"
                     /db_xref="PID:g1117815"
                     /db_xref="GI:1117815"
                     /translation="MVTSVRTQPPVIMPGMQTEIKTQATNLAANLSAVRESATATLSG
                     EIKGPQLEDFPALIKQASLDA"
BASE COUNT      978 a    759 c    925 g    779 t
ORIGIN


alignment_scores:
            Quality:  277.00              Length:     64
              Ratio:  4.328                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-068-804-8 x STU30491   ..

 Align seg 1/1  to: STU30491  from: 1  to: 3441

       1 MetValThrSerValArgThrGlnProProValIleMetProGlyMetGl 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    3248 ATGGTTACAAGTGTAAGGACTCAGCCCCCCGTCATAATGCCAGGTATGCA 3297

      17 nThrGluIleLysThrGlnAlaThrAsnLeuAlaAlaAsnLeuSerAlaV 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    3298 GACCGAGATCAAAACGCAGGCCACGAATCTTGCGGCGAATCTTTCCGCAG 3347

      34 alArgGluSerAlaThrAlaThrLeuSerGlyGluIleLysGlyProGln 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    3348 TCAGAGAAAGTGCCACAGCGACGCTGTCAGGGGAAATTAAAGGCCCGCAA 3397

      51 LeuGluAspPheProAlaLeuIleLysGlnAlaSerLeuAsp 64
         |||||||||||||||||||||||||||||||||||||||||
    3398 CTGGAAGATTTTCCCGCGCTGATCAAACAGGCGAGTCTGGAT 3439

seq_name: gb_ba1:STSIPGNS

seq_documentation_block:
LOCUS       STSIPGNS     5393 bp    DNA             BCT       04-OCT-1995
DEFINITION  S.typhi spaT, sipB, sipC, sipD and sipA genes.
ACCESSION   X82670
NID         g1009001
VERSION     X82670.1  GI:1009001
KEYWORDS    sipA gene; sipB gene; sipC gene; sipD gene; spaT gene.
SOURCE      Salmonella typhi.
  ORGANISM  Salmonella typhi
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Salmonella.
REFERENCE   1  (bases 1 to 5393)
  AUTHORS   Hermant,D., Menard,R., Arricau,N., Parsot,C. and Popoff,M.Y.
  TITLE     Functional conservation of the Salmonella and Shigella effectors of
            entry into epithelial cells
  JOURNAL   Mol. Microbiol. 17 (4), 781-789 (1995)
  MEDLINE   96111497
REFERENCE   2  (bases 1 to 5393)
  AUTHORS   Popoff,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-1994) M. Popoff, Institut Pasteur, Unite des
            Enterobacteries, 28 rue du Docteur Roux, 75724 Paris Cedex 15,
            FRANCE
FEATURES             Location/Qualifiers
     source          1. .5393
                     /organism="Salmonella typhi"
                     /strain="Ty2"
                     /db_xref="taxon:601"
     gene            43. .540
                     /gene="spaT"
     CDS             43. .540
                     /gene="spaT"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA57987.1"
                     /db_xref="PID:g1009002"
                     /db_xref="GI:1009002"
                     /db_xref="SPTREMBL:Q57260"
                     /translation="MDYQNNVSEERVAEMIWDAVSEGATLKDVHGIPQDMMDGLYAHA
                     YEFYNQGRLDEAETFFRFLCIYDFYNPDYTMGLAAVCQLKKQFQKACDLYAVAFTLLK
                     NDYRPVFFTGQCQLLMRKAAKARQCFELVNERTEDESLRAKALVYLEALKTAETEQHS
                     EQEKE"
     gene            543. .2324
                     /gene="sipB"
     CDS             543. .2324
                     /gene="sipB"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA57988.1"
                     /db_xref="PID:g1009003"
                     /db_xref="GI:1009003"
                     /db_xref="SPTREMBL:Q56134"
                     /translation="MVNDASSISRSGYTQNPRLAEAAFEGVRKNTDFLKAADKAFKDV
                     VATKAGDLKAGTKSGESAINTVGLKPPTDAARKLSSEGQLTLLLGKLMTLLGDVSLS
                     QLESRLAVWQAMIESQKEMGIQVSKEFQTALGEAQEATDLYEASIKKTDTAKSVYDAA
                     AKKLTQAQNKLQSLDPADPGYAQAEAAVEQAGKEATEAKEALDKATDATVKAGTDAKA
                     KAEKADNILTKFQGTANAASQNQVSQGEQDNLSNVARLTMLMAMFIEIVGKNTEESLQ
                     NDLALFNALQEGRQAEMEKKSAEFQEETRKAEETNRIMGCIGKVLGALLTIVSVVAAV
                     FTGGASLALAAVGLAVMVADEIVKAATGVSFIQQALNPIMEHVLKPLMELIGKAITKA
                     LEGLGVDKKTAEMAGSIVGAIVAAIAMVAVIVVVAVVGKGAAAKLGNALSKMMGETIK
                     KLVPNVLKQLAQNGSKLFTQGMQRITSGLGNVGSKMGLQTNALSKELVGNTLNKVALG
                     MEVTNTAAQSAGGVAEGVFIKNASEALADFMLARFAMDQIQQWLKQSVEIFGENQKVT
                     AELQKAMSSAVQQNADASRFILRQSRA"
     gene            2352. .3581
                     /gene="sipC"
     CDS             2352. .3581
                     /gene="sipC"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA57989.1"
                     /db_xref="PID:g1009004"
                     /db_xref="GI:1009004"
                     /db_xref="SPTREMBL:Q56135"
                     /translation="MLISNVGINPAAYLNNHSVENSSQTASQSVSAKDILNSIGISSS
                     KVSDLGLSPTLSAPAPGVLTQTPGTITSFLKASIQNTDMNQDLNALANNVTTKANEVV
                     QTQLREQQAEVGKFFDISGMSSSAVALLAAANTLMLTLNQADSKLSGKLSLVSFDAAK
                     TTASSMMREGMNALSGSISQSALQLGITGVGAKLEYKGLQNERGALKHNAAKIDKLTT
```

```
                     ESHSIKNVLNGQNSVKLGAEGVDSLKSLNMKKTGTDATKNLNDATLKSNAGTSATESL
                     GIKNSNKQISPEHQAILSKRLESVESDIRLEQNTMDMTRIDARKMQMTGDLIMKNSVT
                     VGGIAGASRQYAATQERSEQQISQVNNRVASTASDEARESSRKSTSLIQEMLKTMESI
                     NQSKASALAAIAGNIRA"
     gene            3651. .4670
                     /gene="sipD"
     CDS             3651. .4670
                     /gene="sipD"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA57990.1"
                     /db_xref="PID:g1009005"
                     /db_xref="GI:1009005"
                     /db_xref="SPTREMBL:Q56136"
                     /translation="MLNIQNYSASPHPGIVAERPQTPSASEHAEIAVVPSTTEHRGTD
                     IISLSQAATKIQQAQQTLQSTPPISEENNDERTLARQQLTSSLNALAKSGVSLSAEQN
                     ENLRSTFSARRRPYLALRLWPARTTISDAEIWDMVSQNISAIGDSYLGVYENVVAVYT
                     DFYQAFSDILSKMGGWLSPGKDGNTIKLNVDSLKSEISSLINKYTQINKNTILFPSQT
                     GSGMTTATKAEAEQWIKELNLPDSCLKASGSGYVVLVDTGPLSKMVSDLNGIGSGSAL
                     ELDNAKYQAWQSGFKAQEENLKTTLQTLTQKYSNANSLYDNLVKVLSSTISSSLETAK
                     SFLQG"
     gene            4689. .5393
                     /gene="sipA"
     CDS             4689. .>5393
                     /gene="sipA"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA57991.1"
                     /db_xref="PID:g1009006"
                     /db_xref="GI:1009006"
                     /db_xref="SPTREMBL:Q56137"
                     /translation="MVTSVRTQPPVIMPGMQTEIKTQATNLAANLSAVRESATATLSG
                     EIKGQQLEDFPALIKQASLDALFKCGKDAEALKEVFTNSNNVAGKKAIMEFAGLFRSA
                     LNATSDSPEAKTLLMKVGAEYTAQIIKDGLKEKSAFGPWLPETKKAEAKLENLEKQLL
                     DIIKNNTGGELSKLSTNLVMQEVMPYIASCIEHNFGCTLDPLTRSSLTQLVDKAAAKA
                     VEALDMCHQKLTQEQGT"
BASE COUNT     1584 a   1156 c   1458 g   1195 t
ORIGIN


alignment_scores:
            Quality:  271.00              Length:     64
              Ratio:  4.234                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 98.438

alignment_block:
 US-09-068-804-8 x STSIPGNS   ..

 Align seg 1/1  to: STSIPGNS  from: 1  to: 5393

       1 MetValThrSerValArgThrGlnProProValIleMetProGlyMetGl 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    4689 ATGGTTACAAGTGTAAGAACTCAGCCCCCCGTCATAATGCCAGGTATGCA 4738

      17 nThrGluIleLysThrGlnAlaThrAsnLeuAlaAlaAsnLeuSerAlaV 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    4739 GACCGAGATCAAAACGCAGGCCACGAATCTTGCGGCGAATCTTTCCGCGG 4788

      34 alArgGluSerAlaThrAlaThrLeuSerGlyGluIleLysGlyProGln 50
         ||||||||||||||||||||||||||||||||||||||||||||:::|||
    4789 TCAGAGAAAGTGCCACAGCAACGCTGTCAGGGGAAATTAAAGGCCAGCAA 4838

      51 LeuGluAspPheProAlaLeuIleLysGlnAlaSerLeuAsp 64
         ||||||||||||||||||||||||||||||||||||||||||
    4839 CTGGAAGATTTTCCCGCGCTGATCAAACAGGCGAGTCTGGAT 4880

seq_name: gb_ba1:STU40013

seq_documentation_block:
LOCUS       STU40013     3500 bp    DNA             BCT       02-FEB-1996
DEFINITION  Salmonella typhimurium sipD, sipA, and acyl carrier protein (iacP)
            genes, complete cds.
ACCESSION   U40013
```

```
NID         g1172126
VERSION     U40013.1  GI:1172126
KEYWORDS    .
SOURCE      Salmonella typhimurium strain=SL1344.
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Salmonella.
REFERENCE   1  (bases 1 to 3500)
  AUTHORS   Kaniga,K., Trollinger,D. and Galan,J.E.
  TITLE     Identification of two targets of the type III protein secretion
            system encoded by the inv and spa loci of Salmonella typhimurium
            that have homology to the Shigella IpaD and IpaA proteins
  JOURNAL   J. Bacteriol. 177 (24), 7078-7085 (1995)
  MEDLINE   96099289
REFERENCE   2  (bases 1 to 3500)
  AUTHORS   Galan,J.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-NOV-1995) Jorge E. Galan, Molecular Gentics and
            Microbiology, School of Medicine, SUNY Stony Brook, Stony Brook, NY
            11794-5222, USA
FEATURES             Location/Qualifiers
     source          1. .3500
                     /organism="Salmonella typhimurium"
                     /strain="SL1344"
                     /db_xref="taxon:602"
     gene            84. .1130
                     /gene="sipD"
     CDS             84. .1130
                     /gene="sipD"
                     /note="homolog of Shigella spp. IpaD"
                     /codon_start=1
                     /transl_table=11
                     /product="SipD"
                     /protein_id="AAA86617.1"
                     /db_xref="PID:g1172127"
                     /db_xref="GI:1172127"
                     /translation="MLNIQNYSASPHPGIVAERPQTPSASEHVETAVVPSTTEHRGTD
                     IISLSQAATKIHQAQQTLQSTPPISEENNDERTLARQQLTSSLNALAKSGVSLSAEQN
                     ENLRSAFSAPTSALFSASPMAQPRTTISDAEIWDMVSQNISAIGDSYLGVYENVVAVY
                     TDFYQAFSDILSKMGGWLLPGKDGNTVKLDVTSLKNDLNSLVNKYNQINSNTVLFPAQ
                     SGSGVKVATEAEARQWLSELNLPNSCLKSYGSGYVVTVDLTPLQKMVQDIDGLGAPGK
                     DSKLEMDNAKYQAWQAGFKAQEENMKTTLQTLTQKYSNANSLYDNLVKVLSSTISSSW
                     KPPKASCKDNRRGY"
     gene            1133. .3187
                     /gene="sipA"
     CDS             1133. .3187
                     /gene="sipA"
                     /note="homolog of Shigella spp. IpaA"
                     /codon_start=1
                     /transl_table=11
                     /product="SipA"
                     /protein_id="AAA86618.1"
                     /db_xref="PID:g1172128"
                     /db_xref="GI:1172128"
                     /translation="MVTSVRTQPPVIMPGMQTEIKTQATNLAANLSAVRAKCHSDAVR
                     EIKGPQLEDFPALIKQASLDALFKCGKDAEALKEVFTNSNNVAGKKAIMEFAGLFRSA
                     LNATSDSPEAKTLLMKVGAEYAAQIIKDGLKEKSAFGPWLPETKKAEAKLENLEKQLL
                     DIIKNNTGGELSKLSTNLVMQEVMPYIASCIEHNFGCTLDPLTRSNLTHLVDKAAAKA
                     VEALDMCPQKLTQEQGTSVGREARHLEMQTLIPLLLRNVFAQIPADKLPDPKIPEPAA
                     GPVPDGGKKAEPTGININIDSSNHSVDNSKHINNAEPVDNGQRHIDNSNHDNSRKT
                     IDNSRTFIDNSQRNGESHHSTNSSNVSHSHSRVDSTTHQTETAHSASTGAIDHGIAGK
                     IDVTAHATAEAVTNASSESKDGKVVTSEKGTTGETTSFDEVDGVTSKSIIGKPVQATV
                     HGVDDNKQQSQTAEIVNVKPLASQLAGVENVKTDTLQSDTTVITGNKAGTTDNDNSQT
                     DKTGPFSGLKFKQNSFLSTVPSVTNMHSMHFDARETFLGVIRKALEPDTSTPFPVRRA
                     FDGLRAEILPNDTIKSAALKAQCSDIDKHPELKAKMETLKEVITHHPQKEKLAEIALQ
                     FAREAGLTRLKGETDYVLSNVLDGLIGDGSWRAGPAYESYLNKPGVDRVITTVDGLHM
                     QR"
     gene            3206. .3454
                     /gene="iacP"
     CDS             3206. .3454
                     /gene="iacP"
                     /note="IacP"
                     /codon_start=1
```

```
                     /transl_table=11
                     /product="acyl carrier protein"
                     /protein_id="AAA86619.1"
                     /db_xref="PID:g1172129"
                     /db_xref="GI:1172129"
                     /translation="MNMDIEARVKKVITSCIAVDVDSINGQTHLVEDLYADSLDLIDI
                     VFGLSEEFDISCNENDLPDMMTFADICRVVKKSLESRV"
BASE COUNT      1048 a    759 c    879 g    814 t
ORIGIN


alignment_scores:
            Quality:  235.00            Length:     64
              Ratio:  4.123               Gaps:      0
 Percent Similarity:  89.062   Percent Identity: 85.938

alignment_block:
 US-09-068-804-8 x STU40013   ..

 Align seg 1/1  to: STU40013  from: 1  to: 3500

       1 MetValThrSerValArgThrGlnProProValIleMetProGlyMetGl 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1133 ATGGTTACAAGTGTAAGGACTCAGCCCCCCGTCATAATGCCAGGTATGCA 1182

      17 nThrGluIleLysThrGlnAlaThrAsnLeuAlaAlaAsnLeuSerAlaV 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1183 GACCGAGATCAAAACGCAGGCCACGAATCTTGCGGCGAATCTTTCCGCAG 1232

      34 alArgGluSerAlaThrAlaThrLeuSerGlyGluIleLysGlyProGln 50
         |||||:::         :::             ||||||||||||||||||
    1233 TCAGAGCAAAGTGCCACAGCGACGCTGTCAGGGAAATTAAAGGCCCGCAA 1282

      51 LeuGluAspPheProAlaLeuIleLysGlnAlaSerLeuAsp 64
         ||||||||||||||||||||||||||||||||||||||||||
    1283 CTGGAAGATTTTCCCGCGCTGATCAAACAGGCGAGTCTGGAT 1324

seq_name: gb_pl1:YLSEC62

seq_documentation_block:
LOCUS       YLSEC62      1256 bp    DNA             PLN       13-FEB-1997
DEFINITION  Y.lipolytica SEC62 gene.
ACCESSION   X99537
NID         g1465715
VERSION     X99537.1  GI:1465715
KEYWORDS    promoter; SEC62 gene.
SOURCE      Yarrowia lipolytica.
  ORGANISM  Yarrowia lipolytica
            Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
            Dipodascaceae; Yarrowia.
REFERENCE   1  (bases 1 to 1256)
  AUTHORS   Swennen,D., Joyet,P. and Gaillardin,C.
  TITLE     Cloning the Yarrowia lipolytica homologue of the Saccharomyces
            cerevisiae SEC62 gene
  JOURNAL   Curr. Genet. 31 (2), 128-132 (1997)
  MEDLINE   97174109
REFERENCE   2  (bases 1 to 1256)
  AUTHORS   Swennen,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUL-1996) D. Swennen, Institut National Agronomique,
            Laboratoire De Gntique Molculaire Et, Thiverval Grignon, 78850,
            FRANCE
FEATURES             Location/Qualifiers
     source          1. .1256
                     /organism="Yarrowia lipolytica"
                     /strain="W29"
                     /db_xref="taxon:4952"
     gene            15. .1205
                     /gene="SEC62"
     CDS             15. .1205
                     /gene="SEC62"
                     /codon_start=1
                     /protein_id="CAA67878.1"
                     /db_xref="PID:e255564"
                     /db_xref="PID:g1465716"
                     /db_xref="GI:1465716"
                     /db_xref="SWISS-PROT:Q99161"
                     /translation="MTENVPQGQITMPLQQGGAREISPQALAVADYLRSHKLLKQRPG
                     ILNGKRSDFFRVKRAIRALEDPKYKQLQSKPNSKLPPINSRNEAISIFRLMPINQMAL
                     RVDKLPTQTALMMKQKPEQGVPVLQVNPQQEFGDDMYYTWFYNPVPLTTYLYGALGVA
                     AIFAGVLFPLWPIFLRQGVWYLSVGMLGLIGVFFGIALVRLVIFVLTWPTVKPGIWIF
                     PNLFADVGFVDSFIPLWVWHGTPERDLLPQKFKNKKKKKNAGTVIESKEPPRKLTKEE
                     KQKQKEANAQMEQMQAAFQTQLSSFATQMQQIKEMSDSGIDPQIIAAQLQAQYPPDKQ
                     AAIKLENEQAQAKLDERIRELAAQIQDKTNANKTGDKIEEVADKENKEAPKRIVTLED
                     ANDE"
BASE COUNT      332 a    355 c    318 g    251 t
ORIGIN


alignment_scores:
            Quality:   62.00            Length:     52
              Ratio:  1.590               Gaps:      2
 Percent Similarity:  75.000   Percent Identity: 26.923

alignment_block:
 US-09-068-804-8 x YLSEC62   ..

 Align seg 1/1  to: YLSEC62  from: 1  to: 1256

      13 MetProGlyMetGlnThrGluIleLysThrGlnAlaThrAsnLeuAlaAl 29
         |||      ||||||::::::::::::::||||||   :::::::::|||::
     873 ATGGAGCAGATGCAGGCTGCTTTCCAGACCCAGCTCTCCAGCTTCGCCAC 922

      29 aAsnLeuSerAlaValArgGluSerAlaThrAlaThrLeuSerGlyGluI 46
         :::::::      ::::::|||   :::   :::   ::::::   :::|
     923 TCAGATGCAACAGATCAAGGAGATGTCCGACTCTGGCATCGACCCCCAGA 972

      46 leLysGlyProGlnLeuGlu...AspPheProAlaLeuIleLysGlnAla 61
         ||   ::::::||||||:::   ::::::|||:::      |||||||||
     973 TCATTGCTGCCCAGCTCCAGGCTCAGTACCCTCCC...GATAAGCAGGCT 1019

      62 SerLeu 63
         ::::::
    1020 GCTATC 1025

seq_name: gb_pl1:WHTGLUTB

seq_documentation_block:
LOCUS       WHTGLUTB     2999 bp    DNA             PLN       17-DEC-1992
DEFINITION  Tricticum aestivum, high MW glutenin subunit (Bx7) gene, complete
            cds.
ACCESSION   M22209
NID         g170744
VERSION     M22209.1  GI:170744
KEYWORDS    glutenin.
SOURCE      T.aestivum (cv Cheyenne) DNA.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Triticum.
REFERENCE   1  (bases 1 to 2999)
  AUTHORS   Anderson,O.D. and Greene,F.C.
  TITLE     The characterization and comparative analysis of high MW glutenin
            genes from genomes A & B of hexaploid wheat
  JOURNAL   Theor. Appl. Genet. (1989) In press
COMMENT     Computer readable copy of sequence [1] kindly submitted by
            O.D.Anderson 11-JAN-1989.
FEATURES             Location/Qualifiers
     source          1. .2999
                     /organism="Triticum aestivum"
                     /db_xref="taxon:4565"
     CDS             501. .2870
                     /note="high MW glutenin subunit (Bx7)"
                     /codon_start=1
                     /protein_id="AAA34291.1"
                     /db_xref="PID:g170745"
```

```
                     /db_xref="GI:170745"
                     /translation="MAKRLVLFAAVVVALVALTAAEGEASGQLQCEHELEACQQVVDQ
                     QLRDVSPGCRPITVSPGTRQYEQQPVVPSKAGSFYPSETTPSQQLQQMIFWGIPALLR
                     RYYPSVTSSQQGSYYPGQASPQQSGQGQQPGQEQQPGQGQQDQQPGQRQQGYYPTSPQ
                     QPGQGQQLGQGQPGYYPTSQQPGQKQQAGQGQQSGQGQQGYYPTSPQQSGQGQQPGQG
                     QPGYYPTSPQQSGQWQQPGQGQQPGQGQQSGQGQQGQQPGQGQRPGQGQQGYYPISPQ
                     QPGQGQQSGQGQPGYYPTSLRQPGQWQQPGQGQQPGQGQQGQQPGQGQQSGQGQQGYY
                     PTSLQQPGQGQQLGQGQPGYYPTSQQSEQGQQPGQGKQPGQGQQGYYPTSPQQSGQGQ
                     QLGQGQPGYYPTSPQQSGQGQQSGQGQQGYYPTSPQQSGQGQQPGQGQSGYFPTSRQQ
                     SGQGQQPGQGQQSGQGQQGQQPGQGQQAYYPTSSQQSRQRQQAGQWQRPGQGQPGYYP
                     TSPQQPGQEQQSGQAQQSGQWQLVYYPTSPQQPGQLQQPAQGQQPAQGQQSAQEQQPG
                     QAQQSGQWQLVYYPTSPQQPGQLQQPAQGQQGYYPTSPQQSGQGQQGYYPTSPQQSGQ
                     GQQGYYPTSPQQSGQGQQPGQGQQPRQGQQGYYPISPQQSGQGQQPGQGQQGYYPTSP
                     QQSGQGQQPGHEQQPGQWLQPGQGQQGYYPTSSQQSGQGHQSGQGQQGYYPTSLWQPG
                     QGQQGYASPYHVSAEYQAARLKVAKAQQLAAQLPAMCRLEGSDALSTRQ"
BASE COUNT      987 a    873 c    713 g    426 t
ORIGIN      Chromosome 1B.


alignment_scores:
            Quality:   64.00              Length:     61
              Ratio:  1.422                Gaps:      2
 Percent Similarity:  73.770   Percent Identity: 26.230

alignment_block:
 US-09-068-804-8 x WHTGLUTB   ..

 Align seg 1/1  to: WHTGLUTB  from: 1  to: 2999

       3 ThrSerValArgThrGlnProProValIleMetProGlyMetGlnThrGl 19
         |||:::|||||||||:::::::::   ::::::|||:::      :::
    2312 ACAACAGTCAGGACAAGGGCAACAAGGGTACTACCCAACTTCTCCGCAAC 2361

      19 uIleLysThrGlnAlaThrAsnLeuAlaAlaAsnLeuSerAla..... 33
         ::::::|||:::|||||||   :::   :::|||:::|||||||
    2362 AGTCAGGACAAGGGCAACAAGGGTACTACCCAACTTCTCCGCAACAGTCA 2411

      34 ..ValArgGluSerAlaThrAlaThrLeuSerGlyGluIleLysGlyPro 49
           :::|||::::::|||   :::      ::::::   ::::::::::::
    2412 GGACAAGGGCAGCAGCCAGGACAAGGACAACAGCCAAGACAAGGGCAACA 2461

      50 Gln...LeuGluAspPheProAlaLeuIleLys 59
         :::   |||   :::|||:::|||   ::::::
    2462 AGGGTACTACCCAATTTCTCCGCAGCAGTCAGG 2494

seq_name: gb_ba1:PAARCC

seq_documentation_block:
LOCUS       PAARCC       1346 bp    DNA             BCT       12-SEP-1993
DEFINITION  Pseudomonas aeruginosa arcC gene for carbamate kinase (EC 2.7.2.2).
ACCESSION   X14693
NID         g45289
VERSION     X14693.1  GI:45289
KEYWORDS    arcABC operon; arcC gene; carbamate kinase.
SOURCE      Pseudomonas aeruginosa.
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
            Pseudomonas.
REFERENCE   1  (bases 1 to 1346)
  AUTHORS   Baur,H., Luethi,E., Stalon,V., Mercenier,A. and Haas,D.
  TITLE     Sequence analysis and expression of the arginine-deiminase and
            carbamate-kinase genes of Pseudomonas aeruginosa
  JOURNAL   Eur. J. Biochem. 179 (1), 53-60 (1989)
  MEDLINE   89137094
FEATURES             Location/Qualifiers
     source          1. .1346
                     /organism="Pseudomonas aeruginosa"
                     /strain="PAO1"
                     /db_xref="taxon:287"
                     /clone="pME178"
     repeat_unit     59. .70
                     /note="inverted repeat A"
     repeat_unit     76. .87
                     /note="inverted repeat A'"
     RBS             97. .100
                     /note="pot.ribosome binding site"
     CDS             110. .1042
                     /note="carbamate kinase (AA 1-310)"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA32823.1"
                     /db_xref="PID:g45290"
                     /db_xref="GI:45290"
                     /db_xref="SWISS-PROT:P13982"
                     /translation="MRIVVALGGNALLRRGEPMTADNQRENVRIAAEQIAKVAPGNEL
                     VIAHGNGPQVGLLALQGAAYDKVSPYPLDVLGAETEGMIGYMIEQEMGNLLPFEVPFA
                     TILTQVEVDGKDPAFQNPTKPIGPVYSREEAERLAAEKGWSIAPDGDKFRRVVPSPRP
                     KRIFEIRPVKWLLEKGTIVICAGGGGIPTMYDEAGKKLSGVEAVIDKDLCSSLLAQEL
                     VADILIIATDVDAAYVDWGKPTQKAIAQAHPDELERLGFAAGSMGPKVQAAIEFARAT
                     GKDAVIGSLADIVAITEGKAGTRVSTRKAGIEYR"
     terminator      1055. .1076
                     /note="pot.stem-loop structure"
BASE COUNT      256 a    480 c    416 g    194 t
ORIGIN


alignment_scores:
            Quality:   61.00              Length:     56
              Ratio:  1.649                Gaps:      1
 Percent Similarity:  66.071   Percent Identity: 30.357

alignment_block:
 US-09-068-804-8 x PAARCC/rev   ..

 Align seg 1/1  to reverse of: PAARCC  from: 1  to: 1346

       8 GlnProProValIleMetProGlyMetGlnThrGluIleLysThrGlnAl 24
         |||:::|||:::      ::::::::::::      ::::::::::::||
     628 CAGCAGCCACTTCACCGGGCGGATCTCGAAGATCCGCTTCGGCCGCGGGC 579

      24 aThrAsnLeuAlaAlaAsnLeuSerAlaValArgGluSerAlaThrAlaT 41
         |   :::   ||||||:::|||   |||||||||   :::|||:::|||
     578 TGGGCACCACGCGGCGGAACTTGTCGCCGTCCGGGGCGATGCTCCAGCCC 529

      41 hrLeuSerGlyGluIleLysGlyProGlnLeuGluAspPheProAlaLeu 57
           |||:::|||:::               |||      ||||||::::::
     528 TTCTCGGCGGCCAG.............GCGCTCGGCTTCTTCCCTGGAGTA 491

      58 IleLysGlnAlaSerLeu 63
         :::   :::   |||
     490 GACCGGGCCGATCGGCTT 473

seq_name: gb_ov:EMU87434

seq_documentation_block:
LOCUS       EMU87434       439 bp    DNA             VRT       02-APR-1997
DEFINITION  Eleginops maclovinus 28S ribosomal RNA gene, D2 domain.
ACCESSION   U87434
NID         g1916896
VERSION     U87434.1  GI:1916896
KEYWORDS    .
SOURCE      Eleginops maclovinus.
  ORGANISM  Eleginops maclovinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
            Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Percomorpha;
            Perciformes; Notothenioidei; Nototheniidae; Eleginops.
REFERENCE   1  (bases 1 to 439)
  AUTHORS   Lecointre,G., Bonillo,C., Ozouf-Costaz,C. and Hureau,J.C.
  TITLE     Molecular evidence for the origins of Antarctic Fishes: paraphyly
            of the Bovichtidae and no indication for the monophyly of the
            Notothenioidei (Teleostei)
  JOURNAL   Polar Biol. (1997) In press
REFERENCE   2  (bases 1 to 439)
  AUTHORS   Lecointre,G. and Ritchie,P.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JAN-1997) Service de Systematique Moleculaire, Museum
```

```
            National d'Histoire Naturelle, 43 rue Cuvier, Paris 75231 Cedex 05,
            France
FEATURES             Location/Qualifiers
     source          1. .439
                     /organism="Eleginops maclovinus"
                     /db_xref="taxon:56733"
     rRNA            <1. .>439
                     /note="D2 domain"
                     /product="28S ribosomal RNA"
BASE COUNT       42 a    164 c    167 g     66 t
ORIGIN


alignment_scores:
            Quality:   56.50            Length:     48
              Ratio:  1.823              Gaps:      1
 Percent Similarity:  64.583   Percent Identity: 33.333

alignment_block:
 US-09-068-804-8 x EMU87434/rev   ..

 Align seg 1/1  to reverse of: EMU87434  from: 1  to: 439

       6 ArgThrGlnProProValIleMetProGlyMetGlnThrGluIleLysTh 22
         |||:::         ::::::   ||||||   |||::::::
     360 CGGAGCGCACTGAGGACAGTCCGCCCCGGTCGACAGTCGCACCGGGAGCA 311

      22 rGlnAlaThrAsnLeuAlaAlaAsnLeuSerAlaValArgGluSerAlaT 39
           |||::::::|||:::|||   |||::::::   ||||||:::|||
     310 GGGGGCCCCGTCCCTCCCCGCCCCGCTGGGGGGAGGGAGAGAGGGCGCAC 261

      39 hrAlaThrLeu..............SerGlyGluIleLysGly 48
           :::||||||               |||||||||:::::::::
     260 GGAGCACTTTGTCCACGGCCCCGGGAAGCGGCGAGGTCCGGGCG 217

seq_name: gb_vi:PPVNIB

seq_documentation_block:
LOCUS        PPVNIB        366 bp ss-RNA             VRL       30-MAY-1996
DEFINITION   Plum pox virus coat protein (CP) and NIb gene, partial cds.
ACCESSION    L42472
NID          g833744
VERSION      L42472.1  GI:833744
KEYWORDS     coat protein.
SOURCE       Plum pox virus.
  ORGANISM   Plum pox virus
             Viruses; ssRNA positive-strand viruses, no DNA stage; Potyviridae;
             Potyvirus.
REFERENCE    1  (bases 1 to 366)
  AUTHORS    Lopez-Moya,J.J., Canto,T., Diaz-Ruiz,J.R. and Lopez-Abella,D.
  TITLE      Transmission by aphids of a naturally non-transmissible plum pox
             virus isolate with the aid of potato virus Y helper component
  JOURNAL    J. Gen. Virol. 76 (Pt 9), 2293-2297 (1995)
  MEDLINE    96005051
COMMENT      Partial sequence of PPV isolate 3.3.
FEATURES             Location/Qualifiers
     source          1. .366
                     /organism="Plum pox virus"
                     /isolate="3.3"
                     /db_xref="taxon:12211"
     gene            1. .366
                     /gene="NIb"
     mat_peptide     <1. .147
                     /gene="NIb"
     CDS             <1. .>366
                     /gene="NIb"
                     /note="CP gene"
                     /codon_start=1
                     /product="coat protein"
                     /protein_id="AAB00932.1"
                     /db_xref="PID:g833745"
                     /db_xref="GI:833745"
                     /translation="DGKAPYIAETALKKLYTDTEASETEIERYLEAFYDDINDDGESN
                     VVVHQADEREDEEEVDALQPPPVIQPAPRTTAPMLNPIFTPATTQPATKPVSQVSGPQ
                     LQTFGTYGNEDASPSNSNAL"
     mat_peptide     148. .>366
                     /gene="CP"
                     /product="coat protein"
     gene            148. .366
                     /gene="CP"
BASE COUNT       118 a     96 c     80 g     72 t
ORIGIN


alignment_scores:
            Quality:   55.50            Length:     47
              Ratio:  1.734              Gaps:      1
 Percent Similarity:  68.085   Percent Identity: 27.660

alignment_block:
 US-09-068-804-8 x PPVNIB   ..

 Align seg 1/1  to: PPVNIB  from: 1  to: 366

       8 GlnProProValIleMetProGlyMetGlnThrGluIleLysThrGlnAl 24
         |||||||||          |||   ::::|::::::   :::|||   ||
     187 CAACCACCT.........CCAGTCATACAGCCTGCACCCCGGACTACGGC 227

      24 aThrAsnLeuAlaAlaAsnLeuSerAlaValArgGluSerAlaThrAlaT 41
         |:::   |||:::::::   ::::::::::::          ::::::::::
     228 GCCAATGCTCAACCCCATTTTCACGCCAGCAACAACTCAACCAGCAACAA 277

      41 hrLeuSerGlyGluIleLysGlyProGlnLeuGluAspPhe 54
         ::      :::::::::   ||||||||||||:::   |||
     278 AACCAGTTTCACAGGTGTCAGGACCTCAACTGCAAACTTTT 318

seq_name: gb_pl1:TAGLB11B

seq_documentation_block:
LOCUS        TAGLB11B     4034 bp    DNA              PLN       25-JUN-1996
DEFINITION   Wheat Glu-B1-1b gene for HMW glutenin subunit.
ACCESSION    X13927 M22209
NID          g1403096
VERSION      X13927.1  GI:1403096
KEYWORDS     Glu-B1-1b gene; glutenin.
SOURCE       Triticum aestivum.
  ORGANISM   Triticum aestivum
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
             Poaceae; Triticum.
REFERENCE    1  (bases 1 to 4034)
  AUTHORS    Anderson,O.D.
  TITLE      Direct Submission
  JOURNAL    Submitted (07-JAN-1989) Anderson O.D., USDA, Western Regional
             Research Centre, 800 Buchanan Street, Albany, CA 94710, USA
  REMARK     Revised by [3]
REFERENCE    2  (bases 1 to 4034)
  AUTHORS    Anderson,O.D. and Greene,F.C.
  TITLE      The characterization and comparative analysis of high MW glutenin
             genes from genomes A & B of hexaploid wheat
  JOURNAL    Theor. Appl. Genet. 77, 689-700 (1989)
REFERENCE    3  (bases 1 to 4034)
  AUTHORS    Anderson,O.D.
  TITLE      Direct Submission
  JOURNAL    Submitted (10-JUN-1996) Anderson O.D., USDA, Western Regional
             Research Centre, 800 Buchanan Street, Albany, CA 94710, USA
COMMENT      On Jul 5, 1996 this sequence version replaced gi:21748.
             library=lambda EMBL4.
FEATURES             Location/Qualifiers
     source          1. .4034
                     /organism="Triticum aestivum"
                     /cultivar="Cheyenne"
                     /db_xref="taxon:4565"
                     /dev_stage="embryo"
                     /map="1B"
     promoter        1445. .1451
```

```
                     /note="TATA-box"
     CDS             1536. .3905
                     /note="HMW glutenin subunit (AA 1-789)"
                     /codon_start=1
                     /protein_id="CAA32115.1"
                     /db_xref="PID:g21749"
                     /db_xref="GI:21749"
                     /db_xref="SPTREMBL:Q42451"
                     /translation="MAKRLVLFAAVVVALVALTAAEGEASGQLQCEHELEACQQVVDQ
                     QLRDVSPGCRPITVSPGTRQYEQQPVVPSKAGSFYPSETTPSQQLQQMIFWGIPALLR
                     RYYPSVTSSQQGSYYPGQASPQQSGQGQQPGQEQQPGQGQQDQQPGQRQQGYYPTSPQ
                     QPGQGQQLGQGQPGYYPTSQQPGQKQQAGQGQQQSGQGQQGYYPTSPQQSGQGQQPGQG
                     QPGYYPTSPQQSGQWQQPGQGQQPGQGQQSGQGQQGQQPGQGQRPGQGQQGYYPISPQ
                     QPGQGQQSGQGQPGYYPTSLRQPGQWQQPGQGQQPGQGQQGQQPGQGQQSGQGQQGYY
                     PTSLQQPGQGQQLGQGQPGYYPTSQQSEQGQQPGQGKQPGQGQQGYYPTSPQQSGQGQ
                     QLGQGQPGYYPTSPQQSGQGQQSGQGQQGYYPTSPQQSGQGQQPGQGQSGYFPTSRQQ
                     SGQGQQPGQGQQSGQGQQGQQPGQGQQAYYPTSSQQSRQRQQAGQWQRPGQGQPGYYP
                     TSPQQPGQEQQSGQAQQSGQWQLVYYPTSPQQPGQLQQPAQGQQPAQGQQSAQEQQPG
                     QAQQSGQWQLVYYPTSPQQPGQLQQPAQGQQGYYPTSPQQSGQGQQGYYPTSPQQSGQ
                     GQQGYYPTSPQQSGQGQQPGQGQQPRQGQQGYYPISPQQSGQGQQPGQGQQGYYPTSP
                     QQSGQGQQPGHEQQPGQWLQPGQGQQGYYPTSSQQSGQGHQSGQGQQGYYPTSLWQPG
                     QGQQGYASPYHVSAEYQAARLKVAKAQQLAAQLPAMCRLEGSDALSTRQ"
     repeat_region   1842. .3776
     polyA_signal    3960. .3965
BASE COUNT     1348 a   1061 c    898 g    727 t
ORIGIN


alignment_scores:
           Quality:   64.00            Length:     61
             Ratio:  1.422              Gaps:      2
 Percent Similarity:  73.770   Percent Identity: 26.230

alignment_block:
 US-09-068-804-8 x TAGLB11B   ..

 Align seg 1/1  to: TAGLB11B  from: 1  to: 4034

       3 ThrSerValArgThrGlnProProValIleMetProGlyMetGlnThrGl 19
         |||:::|||||||||:::::::::   ::::::|||:::      :::
    3347 ACAACAGTCAGGACAAGGGCAACAAGGGTACTACCCAACTTCTCCGCAAC 3396

      19 uIleLysThrGlnAlaThrAsnLeuAlaAlaAsnLeuSerAla..... 33
         ::::::|||:::||||||   :::   :::|||:::||||||
    3397 AGTCAGGACAAGGGCAACAAGGGTACTACCCAACTTCTCCGCAACAGTCA 3446

      34 ..ValArgGluSerAlaThrAlaThrLeuSerGlyGluIleLysGlyPro 49
           :::|||::::::|||   :::      ::::::   ::::::::::::
    3447 GGACAAGGGCAGCAGCCAGGACAAGGACAACAGCCAAGACAAGGGCAACA 3496

      50 Gln...LeuGluAspPheProAlaLeuIleLys 59
         :::   |||   :::|||:::|||   ::::::
    3497 AGGGTACTACCCAATTTCTCCGCAGCAGTCAGG 3529

seq_name: gb_vi:PPVNIBA

seq_documentation_block:
LOCUS       PPVNIBA       411 bp ss-RNA            VRL       30-MAY-1996
DEFINITION  Plum pox virus coat protein (CP) and NIb gene, partial cds.
ACCESSION   L42473
NID         g833746
VERSION     L42473.1  GI:833746
KEYWORDS    NIb protein; coat protein.
SOURCE      Plum pox virus.
  ORGANISM  Plum pox virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Potyviridae;
            Potyvirus.
REFERENCE   1  (bases 1 to 411)
  AUTHORS   Lopez-Moya,J.J., Canto,T., Diaz-Ruiz,J.R. and Lopez-Abella,D.
  TITLE     Transmission by aphids of a naturally non-transmissible plum pox
            virus isolate with the aid of potato virus Y helper component
  JOURNAL   J. Gen. Virol. 76 (Pt 9), 2293-2297 (1995)
  MEDLINE   96005051
```

```
COMMENT     Partial sequence of PPV isolate 5.15.
FEATURES             Location/Qualifiers
     source          1. .411
                     /organism="Plum pox virus"
                     /isolate="5.15"
                     /db_xref="taxon:12211"
     gene            1. .411
                     /gene="NIb"
     mat_peptide     <1. .147
                     /gene="NIb"
     CDS             <1. .>411
                     /gene="NIb"
                     /note="CP gene"
                     /codon_start=1
                     /protein_id="AAB00933.1"
                     /db_xref="PID:g833747"
                     /db_xref="GI:833747"
                     /translation="DGKAPYIAETALKKLYTDTEASETEIERYLEAFYDDINDDGESN
                     VVVHQADEREDEEEVDAGKPIVATAPAATSPILQPPPVIQPAPRTTAPMLNPIFTSAT
                     TQPATKPVSQVSGPQLQTFGTYGNEDASPSNSNAL"
     mat_peptide     148. .>411
                     /gene="CP"
                     /product="coat protein"
     gene            148. .411
                     /gene="CP"
BASE COUNT      131 a    109 c     91 g     80 t
ORIGIN


alignment_scores:
           Quality:   55.50            Length:     47
             Ratio:  1.734              Gaps:      1
 Percent Similarity:  68.085   Percent Identity: 27.660

alignment_block:
 US-09-068-804-8 x PPVNIBA   ..

 Align seg 1/1  to: PPVNIBA  from: 1  to: 411

       8 GlnProProValIleMetProGlyMetGlnThrGluIleLysThrGlnAl 24
         |||||||||         |||   :::|||::::::   :::|||   ||
     232 CAACCACCT.........CCAGTCATACAGCCTGCACCCCGGACTACGGC 272

      24 aThrAsnLeuAlaAlaAsnLeuSerAlaValArgGluSerAlaThrAlaT 41
         |:::   |||::::::   ::::::::::::         ::::::::::
     273 GCCAATGCTCAACCCCATTTTCACGTCAGCAACAACTCAACCAGCAACAA 322

      41 hrLeuSerGlyGluIleLysGlyProGlnLeuGluAspPhe 54
         ::      :::::::::   ||||||||||||:::   |||
     323 AACCAGTTTCACAGGTGTCAGGACCTCAACTGCAAACTTTT 363

seq_name: gb_in1:LTTRNAGLN

seq_documentation_block:
LOCUS       LTTRNAGLN     696 bp    DNA              INV       28-DEC-1993
DEFINITION  L.tarentolae gene for tRNA gln.
ACCESSION   X68207 S59264
NID         g9665
VERSION     X68207.1  GI:9665
KEYWORDS    transfer RNA; transfer RNA-Gln.
SOURCE      Leishmania tarentolae.
  ORGANISM  Leishmania tarentolae
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Leishmania; lizard Leishmania.
REFERENCE   1  (bases 1 to 696)
  AUTHORS   Suyama,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-SEP-1992) Y. Suyama, Univ of Pennsylvania, Dept of
            Biology, Philadelphia, PA 19104, USA
REFERENCE   2  (bases 1 to 696)
  AUTHORS   Lye,L.F., Chen,D.H. and Suyama,Y.
  TITLE     Selective import of nuclear-encoded tRNAs into mitochondria of the
            protozoan Leishmania tarentolae
```

```
  JOURNAL   Mol. Biochem. Parasitol. 58 (2), 233-245 (1993)
  MEDLINE   93241214
REFERENCE   3  (bases 1 to 696)
  AUTHORS   Sakanyan,V., Desmarez,L., Legrain,C., Charlier,D., Mett,T.,
            Kochikyan,A., Savchenko,A., Boyen,A., Falmagne,P., Pirard,A. and
            Glansdorff,N.
  TITLE     Gene cloning, sequence analysis, purification, and characterization
            of a thermostable aminoacylase from Bacillus stearothermophilus
  JOURNAL   Appl. Environ. Microbiol. 59 (11), 3878-3888 (1993)
  MEDLINE   94113715
FEATURES             Location/Qualifiers
     source          1. .696
                     /organism="Leishmania tarentolae"
                     /strain="UC"
                     /db_xref="taxon:5689"
     tRNA            382. .453
                     /note="codon recognized: CAG"
                     /product="tRNA-Gln"
                     /anticodon=(pos:415. .417,aa:Gln)
BASE COUNT      128 a    135 c    246 g    187 t
ORIGIN


alignment_scores:
            Quality:   56.50            Length:     56
              Ratio:  1.487              Gaps:      1
 Percent Similarity:  67.857   Percent Identity: 23.214

alignment_block:
 US-09-068-804-8 x LTTRNAGLN/rev   ..

 Align seg 1/1  to reverse of: LTTRNAGLN  from: 1  to: 696

       6 ArgThrGlnProProValIleMetProGlyMetGlnThrGluIleLysTh 22
         ::::::   ||||||         |||   ::::::::::::   |||
     637 CAAAGCGCGCCCCCC.........CCCCCCTTACCCCCCGACCGAAAAGG 597

      22 rGlnAlaThrAsnLeuAlaAlaAsnLeuSerAlaValArgGluSerAlaT 39
          ::::::|||   ::::::      ::::::::::::      :::   |
     596 GAAAGTCACCCTAATGACCCGACGAATCCCCACCCTACCCCCCACCCGAA 547

      39 hrAlaThrLeuSerGlyGluIleLysGlyProGlnLeuGluAspPhePro 55
         ||   ::::::   ||||||::::::::::|||   |||   :::||||||
     546 CCTGCAAGGTACATGGCGAGACAAACAACCCCCTCCTACCGAACTTTCCA 497

      56 AlaLeuIleLysGlnAla 61
         :::::::::   |||:::
     496 ACCTTCACACCCCAGCCG 479

seq_name: gb_ba1:AMLAAAR

seq_documentation_block:
LOCUS       AMLAAAR      1400 bp    DNA             BCT       07-FEB-1999
DEFINITION  Amycolatopsis sp. Aaar gene for N-acylamino acid racemase, complete
            cds.
ACCESSION   D30738
NID         g975626
VERSION     D30738.1  GI:975626
KEYWORDS    N-acylamino acid racemase; Aaar.
SOURCE      Amycolatopsis sp. (strain:TS-1-60) DNA, clone:lambda-44.
  ORGANISM  Amycolatopsis sp.
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
            Amycolatopsis.
REFERENCE   1  (bases 1 to 1400)
  AUTHORS   Tokuyama,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-1994) to the DDBJ/EMBL/GenBank databases. Shinji
            Tokuyama, Takeda Chemical Industries, Ltd., Technology Development
            Laboratories.; Mitui Aza Takeda 4720, Hikari, Yamaguchi 743, Japan
            (Tel:0833-71-5575, Fax:0833-71-5645)
REFERENCE   2  (bases 1 to 1400)
  AUTHORS   Tokuyama,S. and Hatano,K.
  TITLE     Cloning, DNA sequencing and heterologous expression of the gene for
            thermostable N-acylamino acid racemase from Amycolatopsis sp.
            TS-1-60 in Escherichia coli
  JOURNAL   Appl. Microbiol. Biotechnol. 42 (6), 884-889 (1995)
  MEDLINE   95226080
FEATURES             Location/Qualifiers
     source          1. .1400
                     /organism="Amycolatopsis sp."
                     /strain="TS-1-60"
                     /db_xref="taxon:37632"
                     /clone="lambda-44"
     gene            62. .1168
                     /gene="Aaar"
     CDS             62. .1168
                     /gene="Aaar"
                     /codon_start=1
                     /transl_table=11
                     /product="N-acylamino acid racemase"
                     /protein_id="BAA06400.1"
                     /db_xref="PID:d1006969"
                     /db_xref="PID:g975627"
                     /db_xref="GI:975627"
                     /translation="MKLSGVELRRVQMPLVAPFRTSFGTQSVRELLLLRAVTPAGEGW
                     GECVTMAGPLYSSEYNDGAEHVLRHYLIPALLAAEDITAAKVTPLLAKFKGHRMAKGA
                     LEMAVLDAELRAHERSFAAELGSVRDSVPCGVSVGIMDTIPQLLDVVGGYLDEGYVRI
                     KLKIEPGWDVEPVRAVRERFGDDVLLQVDANTAYTLGDAPQLARLDPFGLLLIEQPLE
                     EEDVLGHAELARRIQTPICLDESIVSARAAADAIKLGAVQIVNIKPGRVGGYLEARRV
                     HDVCAAHGIPVWCGGMIETGLGRAANVALASLPNFTLPGDTSASDRFYKTDITEPFVL
                     SGGHLPVPTGPGLGVAPIPELLDEVTTAKVWIGS"
BASE COUNT      218 a    464 c    504 g    214 t
ORIGIN


alignment_scores:
            Quality:   59.00            Length:     41
              Ratio:  1.788              Gaps:      1
 Percent Similarity:  80.488   Percent Identity: 31.707

alignment_block:
 US-09-068-804-8 x AMLAAAR   ..

 Align seg 1/1  to: AMLAAAR  from: 1  to: 1400

      13 MetProGlyMetGlnThrGluIleLysThrGlnAlaThrAsnLeuAlaAl 29
         |||:::   :::::::::|||:::::::::::::::   ::::::|||||
     374 ATGGCCGTGCTCGACGCCGAACTCCGCGCGCACGAGAGGTCGTTCGCCGC 423

      29 aAsnLeuSerAlaValArgGluSerAlaThrAlaThrLeuSerGlyGluI 46
         |:::|||::::::||||||:::|||::::::      :::|||      |
     424 CGAACTCGGATCGGTGCGCGATTCTGTGCCGTGCGGCGTTTCGGTCGGGA 473

      46 leLysGly......ProGlnLeu 51
         ||::::::      |||||||||
     474 TCATGGACACCATCCCGCAACTG 496

seq_name: gb_pat:E04131

seq_documentation_block:
LOCUS       E04131       1400 bp    DNA             PAT       29-SEP-1997
DEFINITION  DNA sequence encoding Amycolatopsis acylamino acid recemase.
ACCESSION   E04131
NID         g2172341
VERSION     E04131.1  GI:2172341
KEYWORDS    JP 1992365482-A/1.
SOURCE      Amycolatopsis sp. .
  ORGANISM  Amycolatopsis sp.
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
            Amycolatopsis.
REFERENCE   1  (bases 1 to 1400)
  AUTHORS   Tokuyama,S., Hatano,K., Nakahama,K. and Takahashi,T. .
  TITLE     DNA AND ITS USE
  JOURNAL   Patent: JP 1992365482-A 1 17-DEC-1992;
```

```
            TAKEDA CHEM IND LTD
COMMENT     OS   Amycolatopsis sp.
            PN   JP 1992365482-A/1
            PD   17-DEC-1992
            PF   13-SEP-1991 JP 1991234800
            PR   14-SEP-1990 JP 90P   245257
            PI   TOKUYAMA SHINJI, HATANO KAZUNORI, NAKAHAMA KAZUO, PI
            TAKAHASHI TAKESHI
            PC   C12N9/90,C12N1/21,C12N15/61,(C12N1/21,C12R1:19),(C12N15/61, PC
              C12R1:01);
            CC   strandedness: Double;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
            CC   *source: strain=TS-1-60;
            CC   *source: deposit=FREM BP-3092;
            FH   Key             Location/Qualifiers
            FH
            FT   5'UTR           1. .61
            FT   CDS             62. .1168
            FT                   /product='Amycolatopsis acymino acid  FT
                       racemase'
            FT   3'UTR           1169. .1400.
FEATURES              Location/Qualifiers
     source           1. .1400
                      /organism="Amycolatopsis sp."
                      /db_xref="taxon:37632"
BASE COUNT       218 a    464 c    504 g    214 t
ORIGIN


alignment_scores:
            Quality:   59.00             Length:     41
              Ratio:  1.788               Gaps:      1
 Percent Similarity:  80.488   Percent Identity: 31.707

alignment_block:
 US-09-068-804-8 x E04131   ..

 Align seg 1/1  to: E04131  from: 1  to: 1400

      13 MetProGlyMetGlnThrGluIleLysThrGlnAlaThrAsnLeuAlaAl 29
         |||:::   :::::::::|||:::::::::::::::   ::::::|||||
     374 ATGGCCGTGCTCGACGCCGAACTCCGCGCGCACGAGAGGTCGTTCGCCGC 423

      29 aAsnLeuSerAlaValArgGluSerAlaThrAlaThrLeuSerGlyGluI 46
         |:::|||:::::|||||:::|||:::::      :::|||      |
     424 CGAACTCGGATCGGTGCGCGATTCTGTGCCGTGCGGCGTTTCGGTCGGGA 473

      46 leLysGly......ProGlnLeu 51
         ||:::::      |||||||||
     474 TCATGGACACCATCCCGCAACTG 496

seq_name: gb_pat:I22022

seq_documentation_block:
LOCUS       I22022       1400 bp    DNA             PAT       18-JUN-1996
DEFINITION  Sequence 1 from patent US 5525501.
ACCESSION   I22022
NID         g1602376
VERSION     I22022.1  GI:1602376
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1400)
  AUTHORS   Tokuyama,M., Hatano,K., Nakahama,K. and Takahashi,T.
  TITLE     DNA Fragment encoding acylamino acid racemase
  JOURNAL   Patent: US 5525501-A 1 11-JUN-1996;
FEATURES              Location/Qualifiers
     source           1. .1400
                      /organism="unknown"
BASE COUNT       218 a    464 c    504 g    214 t
ORIGIN


alignment_scores:
            Quality:   59.00             Length:     41
              Ratio:  1.788               Gaps:      1
 Percent Similarity:  80.488   Percent Identity: 31.707

alignment_block:
 US-09-068-804-8 x I22022   ..

 Align seg 1/1  to: I22022  from: 1  to: 1400

      13 MetProGlyMetGlnThrGluIleLysThrGlnAlaThrAsnLeuAlaAl 29
         |||:::   :::::::::|||:::::::::::::::   ::::::|||||
     374 ATGGCCGTGCTCGACGCCGAACTCCGCGCGCACGAGAGGTCGTTCGCCGC 423

      29 aAsnLeuSerAlaValArgGluSerAlaThrAlaThrLeuSerGlyGluI 46
         |:::|||:::::|||||:::|||:::::      :::|||      |
     424 CGAACTCGGATCGGTGCGCGATTCTGTGCCGTGCGGCGTTTCGGTCGGGA 473

      46 leLysGly......ProGlnLeu 51
         ||:::::      |||||||||
     474 TCATGGACACCATCCCGCAACTG 496

seq_name: gb_pl2:AF000939

seq_documentation_block:
LOCUS       AF000939       739 bp    mRNA            PLN       03-JUN-1997
DEFINITION  Hordeum vulgare aleurone ribonuclease mRNA, partial cds.
ACCESSION   AF000939
NID         g2149999
VERSION     AF000939.1  GI:2149999
KEYWORDS    .
SOURCE      barley.
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Hordeum.
REFERENCE   1  (bases 1 to 739)
  AUTHORS   Rogers,J.C. and Rogers,S.W.
  TITLE     Gibberellin-induced Barley Aleurone RNase
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 739)
  AUTHORS   Rogers,J.C. and Rogers,S.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-APR-1997) Biochemistry, University of Missouri, 117
            Schweitzer Hall, Columbia, MO 65211, USA
FEATURES              Location/Qualifiers
     source           1. .739
                      /organism="Hordeum vulgare"
                      /cultivar="Himalaya"
                      /db_xref="taxon:4513"
     CDS              20. .>739
                      /note="BAR-1; expression induced by gibberellins;
                      expressed in aleurone and not in leaf; contains unique 23
                      amino acid insert not found in other ribonucleases"
                      /codon_start=1
                      /product="aleurone ribonuclease"
                      /protein_id="AAB58718.1"
                      /db_xref="PID:g2150000"
                      /db_xref="GI:2150000"
                      /translation="MKLVVLSALLLLSLVAVSSAEEFDFFYLVQQWPGSFCDTKKGCC
                      FPDTGKPATDFGIHGLWPNYAECKTRGELDGALEMVTRRRKKCWPESCNSERLKLWEI
                      RDLVTELDANWPTLACKGGKSFEFWTHEWEKHGTCSNLDQHGYLATALGFKARHNLTS
                      ILADAGIVPSDTETYFLSSIRDAIREGTGFTANLECNRGVDGETQLFQVYQCIDRDGE
                      NLIDCPLPMQGNCKDRVQLPAF"
BASE COUNT       144 a    237 c    240 g    118 t
ORIGIN


alignment_scores:
            Quality:   56.50             Length:     61
```

```
            Ratio:  1.527                Gaps:      3
 Percent Similarity:  60.656   Percent Identity: 31.148

alignment_block:
 US-09-068-804-8 x AF000939/rev   ..

 Align seg 1/1  to reverse of: AF000939  from: 1  to: 739

      6 ArgThrGlnProProVal.................IleMetProGlyMe 16
        |||   :::|||   |||                  :::::::::|||::
    526 CGGCACGATCCCGGCGTCGGCGAGGATGCTGGTGAGGTTGTGGCGGGCCT 477

     16 tGlnThrGluIleLysThrGlnAlaThrAsnLeuAlaAlaAsnLeuSerA 33
        :::::::|||   :::   |||::::::   |||
    476 TGAAGCCGAGCGCCGTCGCCAAGTAGCCGTGCTG.............. 443

     33 laValArgGluSerAlaThrAlaThrLeuSerGlyGluIleLysGlyPro 49
          |||:::   :::|||   |||:::|||         :::   ||||||
    442 ..GTCCAGGTTGGAGCAGGTGCCGTGCTTCTCCCACTCGTG...GGTCCA 398

     50 GlnLeuGluAspPheProAlaLeuIleLysGln 60
        :::||||||::::::|||||||||   :::|||
    397 GAACTCGAAGCTCTTCCCGCCCTTGCACGCCAG 365
```

```
OM of: US-09-068-804-14 to: Issued_Patents_NA:*   out_format : pfs

Date: Oct 5, 1999  7:12 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=x1p
 -Q=/cgn2_1/USPTO_spool/US09068804/runat_05101999_145135_10326/app_query.fasta.1
 -DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=pam150
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=escore
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09068804
 -NCPU=6 -ICPU=3 -WAIT -THREADS=1

Search information block:
Query: US-09-068-804-14
Query length: 513
Database: Issued_Patents_NA:*
Database sequences: 192659
Database length: 52021692
Search time (sec): 101.650000



score_list:
Sequence              Strd Orig    ZScore    EScore Len  ! Documentation  ..
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-817-852A-1 +  152.50  281.08  2.9e-08   906  !
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-470-137-1 +  152.50  281.08  2.9e-08   906  !  S
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-442-063A-47 +  139.50  254.50  8.7e-07   924  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-442-063A-26 +  139.50  253.52  9.9e-07  1002  !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-272-919-1 +  139.50  253.24  1.0e-06  1026  !
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-619-916-1 +  139.50  253.24  1.0e-06  1026  !
/cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:PCT-US95-08542-1 +  139.50  253.24  1.0e-06  1026
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-442-063A-44 +  135.50  247.41  2.2e-06   849  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-442-063A-41 +  127.00  232.32  1.5e-05   711  !
/cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:PCT-US94-07644A-1 +  123.50  196.96  0.0014   745
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-365-486A-6 +   95.00  180.54  0.0115   240  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-452-531-2 +  100.50  178.62  0.0147   711  !  S
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-460-746A-2 +  100.50  178.62  0.0147   711  !
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-460-555-2 +  100.50  178.62  0.0147   711  !  S
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-666-271-4 +  107.00  174.96  0.0235  2880  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-279-058B-39 +   95.00  174.07  0.0263   411  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-442-063A-35 +   95.00  173.64  0.0278   426  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-452-531-1 +  100.50  173.47  0.0284  1091  !
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-460-746A-1 +  100.50  173.47  0.0284  1091  !
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-460-555-1 +  100.50  173.47  0.0284  1091  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-442-063A-38 +   96.50  173.24  0.0293   567  !
/cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:PCT-US95-10509-1 +  104.50  172.91  0.0305  2241
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-666-271-1 +  107.00  171.30  0.0376  3905  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-205-719-1 +   99.50  171.04  0.0388  1128  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-07-949-516A-3 +   98.50  168.19  0.0560  1208  !
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-814-459-3 +   98.50  168.19  0.0560  1208  !
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-821-119-16 +  102.50  167.92  0.0580  2424  !
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-483-695-6 +   96.00  166.10  0.0731   943  !  S
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-07-965-285-6 +   96.00  166.10  0.0731   943  !  S
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-487-231-6 +   96.00  166.10  0.0731   943  !  S
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-483-695-43 +   94.00  162.05  0.1230   943  !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-07-965-285-43 +   94.00  162.05  0.1230   943  !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-487-231-43 +   94.00  162.05  0.1230   943  !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-225-473-2 +   92.50  158.08  0.2047  1019  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-466-603-1 +   92.50  157.69  0.2150  1052  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-314-503A-1 +   92.50  157.69  0.2150  1052  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-468-066-1 +   92.50  157.69  0.2150  1052  !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-466-717-1 +   92.50  157.69  0.2150  1052  !
/cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:PCT-US95-12414-1 +   92.50  157.69  0.2150  1052
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-463-911-6 +   93.00  156.04  0.2657  1313  !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-481-658B-1 +   93.50  155.28  0.2931  1522  !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-567-375-3 +   98.00  153.01  0.3920  3921  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-466-603-3 +   88.50  150.44  0.5449   980  !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-314-503A-3 +   88.50  150.44  0.5449   980
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-468-066-3 +   88.50  150.44  0.5449   980  !

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-817-852A-1

seq_documentation_block:
; Sequence 1, Application US/07817852A
; Patent No. 5427939
;  GENERAL INFORMATION:
;    APPLICANT:  Ruggeri, Zaverio M.
;    APPLICANT:  Ware, Jerry, inventors
;    APPLICANT:  on behalf of The Scripps Research
;    APPLICANT:  Institute
;    TITLE OF INVENTION:  Platelet Transfusion
;    NUMBER OF SEQUENCES:  1
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  The Scripps Research
;      ADDRESSEE:  Institute
;      STREET:  10666 No. 5427939th Torrey Pines Road
;      CITY:  La Jolla
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette-3.5 inch, 1.44 Mb
;      COMPUTER:  AST Bravo IBM PC comp. (386SX)
;      OPERATING SYSTEM:  MS DOS version 3.2
;      SOFTWARE:  WordPerfect 5.1 conv. to ASCII
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/817,852A
;      FILING DATE:  6-Jan-1992
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Barron, Alexis
;      REGISTRATION NUMBER:  22,702
;      REFERENCE/DOCKET NUMBER:  P17,934 USA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 923-4466
;      TELEFAX:  (215) 923-2189
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  906
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single stranded
;      TOPOLOGY:  Linear
US-07-817-852A-1


alignment_scores:
            Quality:  152.50              Length:    230
              Ratio:  1.292                 Gaps:      8
 Percent Similarity:  51.304   Percent Identity: 28.261

alignment_block:
 US-09-068-804-14 x US-07-817-852A-1   ..

 Align seg 1/1  to: US-07-817-852A-1  from: 1  to: 906

      19 LeuThrThrLeuProAspArgLeuProProHisIleThrThrLeuValIl 35
         ||||||:::||||||      ||||||   :::::::|||:::|||   ::
      64 CTGACAGCGCTGCCTCCAGACCTGCCGAAAGACACAACCATCCTCCACCT 113

      35 eProAspAsnAsnLeu.............................T 41
         :::::::|||   |||                             |
     114 GAGTGAGAACCTCCTGTACACCTTCTCCCTGGCAACCCTGATGCCTTACA 163

      41 hrSerLeuProGluLeuProGluGly.........LeuArgGluLeuGlu 54
         ||   |||::::::|||      :::         |||      |||:::
     164 CTCGCCTCACTCAGCTGAACCTAGATAGGTGCGAGCTCACCAAGCTCCAG 213

      55 ValSerGlyAsnLeuGlnLeuThrSerLeuProSerLeuProGlnGlyLe 71
         |||:::|||:::|||
```

```
    214 GTCGATGGGACGCTG..................................... 228

     71 uGlnLysLeuTrpAlaTyrAsnAsnTrpLeuAlaSerLeuProThrLeuP 88

    228 ..................................................... 228

     88 roProGlyLeuGlyAspLeuAlaValSerAsnAsnGlnLeuThrSerLeu 104
          |||   ||||||   |||:::::|||:::|||||||||   ||||||
    229 ..CCAGTGCTGGGGACCCTGGATCTATCCCACAATCAGCTGCAAAGCCTG 276

    105 ProGluMetProProAlaLeuArgGluLeu..................... 114
        |||   :::   :::::::|||   :::|||
    277 CCCTTGCTAGGGCAGACACTGCCTGCTCTCACCGTCCTGGACGTCTCCTT 326

    115 ....ArgValSerGlyAsnAsnLeuThrSerLeuArgAlaLeuProSerG 130
            |||::::::::::      |||   :::|||||||:::|||   :::
    327 CAACCGGCTGACCTCGCTGCCTCTTGGTGCCCTGCGTGGTCTT...GGCG 373

    130 lyLeuGlnLysLeuTrpAlaTyrAsnAsnArgLeuThrSerLeuProGlu 146
          ||||||   |||          :::|||   |||:::::::||||||
    374 AACTCCAAGAGCTCTACCTGAAAGGCAATGAGCTGAAGACCCTGCCCCCA 423

    147 ............MetSerProGlyLeuGlnGluLeuAspValSerHisAs 159
                      :::|||   |||:::   |||:::::::::::::||
    424 GGGCTCCTGATGCCCACACCCAAGCTGGAGAAGCTCAGTCTGGCTAACAA 473

    159 nGlnLeuThrArgLeuProGlnSer...LeuThrGlyLeuSerSerAlaA 175
        |:::||||||   ||||||   :::   |||:::||||||   :::   :
    474 CAACTTGACTGAGCTCCCCGCTGGGCTCCTGAATGGGCTGGAGAATCTCG 523

    175 laArgValTyrLeuAspGlyAsnProLeuSerValArgThrArgAspArg 191
        ::   :::   |||:::   |||:::|||   :::   ::::::::
    524 ACACCCTTCTCCTCCAAGAGAACTCGCTGTATACAATACCAAAGGGCTTT 573

    192 LeuCysGlyHisHisTrpProPheArgHisGlnAspThrLeuArgTyrGl 208
        :::   :::|||      ||||||
    574 TTTGGGTCCCACCTCCTGCCTTTTGC..... 599

    208 yGlyAlaPheArgProArgGluAlaArgAlaLeuHisLeu 221
              |||   |||||||||:::   :::::::::|||
    600 .......TTTCTCCACGGGAACCCCTGGTTATGCAACTG 632

seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-470-137-1

seq_documentation_block:
; Sequence 1, Application US/08470137
; Patent No. 5798216
;  GENERAL INFORMATION:
;    APPLICANT:  Ruggeri, Zaverio M.
;    APPLICANT:  Ware, Jerry, inventors
;    APPLICANT:  on behalf of The Scripps Research
;    APPLICANT:  Institute
;    TITLE OF INVENTION:  Platelet Transfusion
;    NUMBER OF SEQUENCES:  1
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  The Scripps Research
;      ADDRESSEE:  Institute
;      STREET:  10666 No. 5798216th Torrey Pines Road
;      CITY:  La Jolla
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette-3.5 inch, 1.44 Mb
;      COMPUTER:  AST Bravo IBM PC comp. (386SX)
;      OPERATING SYSTEM:  MS DOS version 3.2
;      SOFTWARE:  WordPerfect 5.1 conv. to ASCII
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/470,137
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Barron, Alexis
;      REGISTRATION NUMBER:  22,702
;      REFERENCE/DOCKET NUMBER:  P17,934-A USA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 923-4466
;      TELEFAX:  (215) 923-2189
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  906
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single stranded
;      TOPOLOGY:  Linear
US-08-470-137-1


alignment_scores:
            Quality:  152.50              Length:    230
              Ratio:  1.292                 Gaps:      8
 Percent Similarity:  51.304   Percent Identity: 28.261

alignment_block:
 US-09-068-804-14 x US-08-470-137-1   ..

 Align seg 1/1  to: US-08-470-137-1  from: 1  to: 906

     19 LeuThrThrLeuProAspArgLeuProProHisIleThrThrLeuValIl 35
        ||||||:::||||||      ||||||   :::::::|||:::|||   ::
     64 CTGACAGCGCTGCCTCCAGACCTGCCGAAAGACACAACCATCCTCCACCT 113

     35 eProAspAsnAsnLeu.................................T 41
        ::::::::|||   |||                                 |
    114 GAGTGAGAACCTCCTGTACACCTTCTCCCTGGCAACCCTGATGCCTTACA 163

     41 hrSerLeuProGluLeuProGluGly.........LeuArgGluLeuGlu 54
        ||   |||:::::::|||      :::         |||      |||:::
    164 CTCGCCTCACTCAGCTGAACCTAGATAGGTGCGAGCTCACCAAGCTCCAG 213

     55 ValSerGlyAsnLeuGlnLeuThrSerLeuProSerLeuProGlnGlyLe 71
        |||:::|||:::|||
    214 GTCGATGGGACGCTG..................................... 228

     71 uGlnLysLeuTrpAlaTyrAsnAsnTrpLeuAlaSerLeuProThrLeuP 88

    228 ..................................................... 228

     88 roProGlyLeuGlyAspLeuAlaValSerAsnAsnGlnLeuThrSerLeu 104
          |||   ||||||   |||:::::|||:::|||||||||   ||||||
    229 ..CCAGTGCTGGGGACCCTGGATCTATCCCACAATCAGCTGCAAAGCCTG 276

    105 ProGluMetProProAlaLeuArgGluLeu..................... 114
        |||   :::   :::::::|||   :::|||
    277 CCCTTGCTAGGGCAGACACTGCCTGCTCTCACCGTCCTGGACGTCTCCTT 326

    115 ....ArgValSerGlyAsnAsnLeuThrSerLeuArgAlaLeuProSerG 130
            |||::::::::::      |||   :::|||||||:::|||   :::
    327 CAACCGGCTGACCTCGCTGCCTCTTGGTGCCCTGCGTGGTCTT...GGCG 373

    130 lyLeuGlnLysLeuTrpAlaTyrAsnAsnArgLeuThrSerLeuProGlu 146
          ||||||   |||          :::|||   |||:::::::||||||
    374 AACTCCAAGAGCTCTACCTGAAAGGCAATGAGCTGAAGACCCTGCCCCCA 423

    147 ............MetSerProGlyLeuGlnGluLeuAspValSerHisAs 159
                      :::|||   |||:::   |||:::::::::::::||
    424 GGGCTCCTGATGCCCACACCCAAGCTGGAGAAGCTCAGTCTGGCTAACAA 473

    159 nGlnLeuThrArgLeuProGlnSer...LeuThrGlyLeuSerSerAlaA 175
        |:::||||||   ||||||   :::   |||:::||||||   :::   :
    474 CAACTTGACTGAGCTCCCCGCTGGGCTCCTGAATGGGCTGGAGAATCTCG 523

    175 laArgValTyrLeuAspGlyAsnProLeuSerValArgThrArgAspArg 191
        ::   :::   |||:::   |||:::|||   :::   ::::::::
    524 ACACCCTTCTCCTCCAAGAGAACTCGCTGTATACAATACCAAAGGGCTTT 573

    192 LeuCysGlyHisHisTrpProPheArgHisGlnAspThrLeuArgTyrGl 208
```

```
      :::   :::|||      ||||||
    574 TTTGGGTCCCACCTCCTGCCTTTTGC..................... 599

    208 yGlyAlaPheArgProArgGluAlaArgAlaLeuHisLeu 221
              |||   |||||||||:::   ::::::::::|||
    600 .......TTTTCTCCACGGGAACCCCTGGTTATGCAACTG 632

seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-442-063A-47

seq_documentation_block:
; Sequence 47, Application US/08442063A
; Patent No. 5705609
;  GENERAL INFORMATION:
;    APPLICANT:  RUOSLAHTI, ERKKI I.
;    APPLICANT:  PIERSCHBACHER, MICHAEL D.
;    APPLICANT:  CARDENAS, JOSE
;    APPLICANT:  CRAIG, WILLIAM
;    APPLICANT:  MULLEN, DANIEL G.
;    TITLE OF INVENTION:  DECORIN FRAGMENTS AND METHODS OF
;    TITLE OF INVENTION:  INHIBITING CELL REGULATORY FACTORS
;    NUMBER OF SEQUENCES:  62
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CAMPBELL AND FLORES
;      STREET:  4370 LA JOLLA VILLAGE DRIVE, SUITE 700
;      CITY:  SAN DIEGO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/442,063A
;      FILING DATE:  16-MAY-1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/865,652
;      FILING DATE:  03-APR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CAMPBELL, CATHRYN A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-LA 1454
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-535-9001
;      TELEFAX:  619-535-8949
;  INFORMATION FOR SEQ ID NO:  47:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  924 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(1..915, 919..924)
US-08-442-063A-47


alignment_scores:
            Quality:  139.50              Length:    260
              Ratio:  0.962                Gaps:     14
 Percent Similarity:  55.769   Percent Identity: 25.769

alignment_block:
 US-09-068-804-14 x US-08-442-063A-47   ..

 Align seg 1/1  to: US-08-442-063A-47  from: 1  to: 924

     11 ValLeuAsnValGlyAlaSerGlyLeuThrThrLeuProAspArgLeuPr 27
        |||::::::   ::::::   ||||||   :::::|||      |||||
    106 GTGGTCCAGTGTTCTGATTTGGGTCTGGACAAAGTGCCAAAGGATCTTCC 155

     27 oProHisIleThrThrLeuValIleProAspAsnAsnLeuThrSerLeuP 44
        ||||::::::|||   |||   :::::::::|||:::::::|||
    156 CCCTGACACAACTCTGCTAGACCTGCAAAACAACAAAATAACC....... 198

     44 roGluLeuProGluGly.....LeuArgGluLeuGlu..... 54
          |||:::   :::|||           |||:::::::||:::
    199 ..GAAATCAAAGATGGAGACTTTAAGAACCTGAAGAACCTTCACGCATTG 246

     55 .....ValSerGlyAsnLeu..... 59
             |||:::::::::::::
    247 ATTCTTGTCAACAATAAAATTAGCAAAGTTAGTCCTGGAGCATTTACACC 296

     60 .....GlnLeuThrSerL 64
                        ||||||:::   |
    297 TTTGGTGAAGTTGGAACGACTTTATCTGTCCAAGAATCAGCTGAAGGAAT 346

     64 euPro...SerLeuProGlnGlyLeuGlnLysLeuTrpAlaTyrAsnAsn 79
        |||||      :::|||:::   ||||||   |||:::|||:::::|||
    347 TGCCAGAAAAAATGCCCAAAACTCTTCAGGAGCTGCGTGCCCATGAGAAT 396

     80 TrpLeuAlaSerLeuProThrLeuPro...ProGlyLeuGlyAspLeuAl 95
          ::::::   :::   :::::::::      ||||||:::::::::
    397 GAGATCACCAAAGTGCGAAAAGTTACTTTCAATGGACTGAACCAGATGAT 446

     95 aVal..... 96
         |||
    447 TGTCATAGAACTGGGCACCAATCCGCTGAAGAGCTCAGGAATTGAAAATG 496

     97 .....SerAsnAsn 99
                        :::::::::
    497 GGGCTTTCCAGGGAATGAAGAAGCTCTCCTACATCCGCATTGCTGATACC 546

    100 GlnLeuThrSerLeuProGlu...MetProProAlaLeuArgGluLeuAr 115
        ::::::|||||:::|||:::   :::|||||:::|||   ||||||::
    547 AATATCACCAGCATTCCTCAAGGTCTTCCTCCTTCCCTTACGGAATTACA 596

    115 gValSerGlyAsnAsnLeu.....ThrSerLeuArgAlaL 127
        ::::::::|||||::::::                 :::|||||:::::|
    597 TCTTGATGGCAACAAAATCAGCAGAGTTGATGCAGCTAGCCTGAAAGGAC 646

    127 euProSerGlyLeuGlnLysLeu...TrpAlaTyrAsnAsnArgLeuThr 142
        ||   ::::::|||   ||||||      ::::::|||:::      :::
    647 TG...AATAATTTGGCTAAGTTGGGATTGAGTTTCAACAGCATCTCTGCT 693

    143 .....SerLeuProGluMetSerProGlyLeuGlnGluLeuAs 155
                |||||||::::::   :::|||   |||:::||||||::
    694 GTTGACAATGGCTCTCTGGCCAAC...ACGCCTCATCTGAGGGAGCTTCA 740

    155 pValSerHisAsnGlnLeuThrArgLeuProGlnSerLeuThrGlyLeuS 172
        ::::::::::|||:::||||||||||:::|||   :::|||:::
    741 CTTGGACAACAACAAGCTTACCAGAGTACCTGGTGGGCTGGCAGAGCATA 790

    172 erSerAlaAlaArgValTyrLeuAspGlyAsnProLeuSerValArgThr 188
                  ||||||||||::::::|||   :::||||||
    791 AGTACATCCAGGTTGTCTACCTTCATAACAACAATATCTCTGTAGTTGGA 840

    189 ArgAspArgLeuCys......GlyHisHis 196
          :::   :::|||      ||||||:::
    841 TCAAGTGACTTCTGCCCACCTGGACACAAC 870

seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-442-063A-26

seq_documentation_block:
; Sequence 26, Application US/08442063A
; Patent No. 5705609
;  GENERAL INFORMATION:
;    APPLICANT:  RUOSLAHTI, ERKKI I.
;    APPLICANT:  PIERSCHBACHER, MICHAEL D.
;    APPLICANT:  CARDENAS, JOSE
;    APPLICANT:  CRAIG, WILLIAM
;    APPLICANT:  MULLEN, DANIEL G.
;    TITLE OF INVENTION:  DECORIN FRAGMENTS AND METHODS OF
```

```
;    TITLE OF INVENTION:  INHIBITING CELL REGULATORY FACTORS
;    NUMBER OF SEQUENCES:  62
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CAMPBELL AND FLORES
;      STREET:  4370 LA JOLLA VILLAGE DRIVE, SUITE 700
;      CITY:  SAN DIEGO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/442,063A
;      FILING DATE:  16-MAY-1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/865,652
;      FILING DATE:  03-APR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CAMPBELL, CATHRYN A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-LA 1454
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-535-9001
;      TELEFAX:  619-535-8949
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1002 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(1..993, 997..1002)
US-08-442-063A-26


alignment_scores:
            Quality:  139.50              Length:    260
              Ratio:  0.962                Gaps:     14
 Percent Similarity:  55.769   Percent Identity: 25.769

alignment_block:
 US-09-068-804-14 x US-08-442-063A-26   ..

 Align seg 1/1  to: US-08-442-063A-26  from: 1  to: 1002

      11 ValLeuAsnValGlyAlaSerGlyLeuThrThrLeuProAspArgLeuPr 27
         |||::::::   ::::::   ||||||   :::::::|||      |||||
     106 GTGGTCCAGTGTTCTGATTTGGGTCTGGACAAAGTGCCAAAGGATCTTCC 155

      27 oProHisIleThrThrLeuValIleProAspAsnAsnLeuThrSerLeuP 44
         ||||:::::::|||   |||   :::::::::|||:::::::|||
     156 CCCTGACACAACTCTGCTAGACCTGCAAAACAACAAAATAACC..... 198

      44 roGluLeuProGluGly.............LeuArgGluLeuGlu..... 54
            |||:::   :::|||           |||:::::::|||:::
     199 ..GAAATCAAAGATGGAGACTTTAAGAACCTGAAGAACCTTCACGCATTG 246

      55 ......ValSerGlyAsnLeu........................... 59
              |||:::::::::::::
     247 ATTCTTGTCAACAATAAAATTAGCAAAGTTAGTCCTGGAGCATTTACACC 296

      60 ....................................GlnLeuThrSerL 64
                                            ||||||:::   |
     297 TTTGGTGAAGTTGGAACGACTTTATCTGTCCAAGAATCAGCTGAAGGAAT 346

      64 euPro...SerLeuProGlnGlyLeuGlnLysLeuTrpAlaTyrAsnAsn 79
         |||||      :::|||:::   ||||||   |||:::|||::::::|||

     347 TGCCAGAAAAAATGCCCAAAACTCTTCAGGAGCTGCGTGCCCATGAGAAT 396

      80 TrpLeuAlaSerLeuProThrLeuPro...ProGlyLeuGlyAspLeuAl 95
            ::::::   :::   :::::::::      ||||||::::::::::
     397 GAGATCACCAAAGTGCGAAAAGTTACTTTCAATGGACTGAACCAGATGAT 446

      95 aVal..................................................... 96
          |||
     447 TGTCATAGAACTGGGCACCAATCCGCTGAAGAGCTCAGGAATTGAAAATG 496

      97 ..........................................SerAsnAsn 99
                                                  :::::::::
     497 GGGCTTTCCAGGGAATGAAGAAGCTCTCCTACATCCGCATTGCTGATACC 546

     100 GlnLeuThrSerLeuProGlu...MetProProAlaLeuArgGluLeuAr 115
         ::::::||||||:::|||:::   :::||||||:::|||   |||||||::
     547 AATATCACCAGCATTCCTCAAGGTCTTCCTCCTTCCCTTACGGAATTACA 596

     115 gValSerGlyAsnAsnLeu..............ThrSerLeuArgAlaL 127
         :::::::||||||::::::                ::::||||||::::::|
     597 TCTTGATGGCAACAAAATCAGCAGAGTTGATGCAGCTAGCCTGAAAGGAC 646

     127 euProSerGlyLeuGlnLysLeu...TrpAlaTyrAsnAsnArgLeuThr 142
         ||   :::::::|||   ||||||      ::::::|||:::      :::
     647 TG...AATAATTTGGCTAAGTTGGGATTGAGTTTCAACAGCATCTCTGCT 693

     143 ............SerLeuProGluMetSerProGlyLeuGlnGluLeuAs 155
                     ||||||::::::   :::|||   |||:::||||||::
     694 GTTGACAATGGCTCTCTGGCCAAC...ACGCCTCATCTGAGGGAGCTTCA 740

     155 pValSerHisAsnGlnLeuThrArgLeuProGlnSerLeuThrGlyLeuS 172
         ::::::::::|||:::||||||||||:::|||   :::|||:::
     741 CTTGGACAACAACAAGCTTACCAGAGTACCTGGTGGGCTGGCAGAGCATA 790

     172 erSerAlaAlaArgValTyrLeuAspGlyAsnProLeuSerValArgThr 188
                     |||||||||::::::|||   :::||||||
     791 AGTACATCCAGGTTGTCTACCTTCATAACAACAATATCTCTGTAGTTGGA 840

     189 ArgAspArgLeuCys......GlyHisHis 196
             :::   :::|||      |||||||:::
     841 TCAAGTGACTTCTGCCCACCTGGACACAAC 870

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-272-919-1

seq_documentation_block:
; Sequence 1, Application US/08272919
; Patent No. 5567807
;  GENERAL INFORMATION:
;    APPLICANT:  Craig, William S.
;    APPLICANT:  Harper, John R.
;    APPLICANT:  Hernandez, Sam D.
;    APPLICANT:  Kostel, Paul J.
;    APPLICANT:  Parker, Jonathan R.
;    APPLICANT:  Vedvick, Thomas S.
;    TITLE OF INVENTION:  Processes for the Purification of Human
;    TITLE OF INVENTION:  Recombinant Decorin and the Detection of Guanidinium Ions
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell and Flores
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/272,919
;      FILING DATE:  08-JUL-1994
;      CLASSIFICATION:  435
```

```
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-LA 1040
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1026 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1026
US-08-272-919-1


alignment_scores:
            Quality:  139.50              Length:    260
              Ratio:  0.962                Gaps:     14
 Percent Similarity:  55.769   Percent Identity: 25.769

alignment_block:
 US-09-068-804-14 x US-08-272-919-1   ..

 Align seg 1/1  to: US-08-272-919-1  from: 1  to: 1026

      11 ValLeuAsnValGlyAlaSerGlyLeuThrThrLeuProAspArgLeuPr 27
         |||::::::   ::::::   ||||||   :::::::|||      |||||
     142 GTGGTCCAGTGTTCTGATTTGGGTCTGGACAAAGTGCCAAAGGATCTTCC 191

      27 oProHisIleThrThrLeuValIleProAspAsnAsnLeuThrSerLeuP 44
         ||||::::::|||   |||   ::::::::::|||::::::|||
     192 CCCTGACACAACTCTGCTAGACCTGCAAAACAACAAAATAACC....... 234

      44 roGluLeuProGluGly.............LeuArgGluLeuGlu..... 54
           |||:::   :::|||             |||::::::|||:::
     235 ..GAAATCAAAGATGGAGACTTTAAGAACCTGAAGAACCTTCACGCATTG 282

      55 ......ValSerGlyAsnLeu........................... 59
               |||:::::::::::::
     283 ATTCTTGTCAACAATAAAATTAGCAAAGTTAGTCCTGGAGCATTTACACC 332

      60 ...................................GlnLeuThrSerL 64
                                           ||||||:::   |
     333 TTTGGTGAAGTTGGAACGACTTTATCTGTCCAAGAATCAGCTGAAGGAAT 382

      64 euPro...SerLeuProGlnGlyLeuGlnLysLeuTrpAlaTyrAsnAsn 79
         |||||      :::|||:::   ||||||   |||:::|||::::::|||
     383 TGCCAGAAAAAATGCCCAAAACTCTTCAGGAGCTGCGTGCCCATGAGAAT 432

      80 TrpLeuAlaSerLeuProThrLeuPro...ProGlyLeuGlyAspLeuAl 95
            ::::::   :::   :::::::::      ||||||::::::::::
     433 GAGATCACCAAAGTGCGAAAAGTTACTTTCAATGGACTGAACCAGATGAT 482

      95 aVal..................................................... 96
           |||
     483 TGTCATAGAACTGGGCACCAATCCGCTGAAGAGCTCAGGAATTGAAAATG 532

      97 ..........................................SerAsnAsn 99
                                                 :::::::::
     533 GGGCTTTCCAGGGAATGAAGAAGCTCTCCTACATCCGCATTGCTGATACC 582

     100 GlnLeuThrSerLeuProGlu...MetProProAlaLeuArgGluLeuAr 115
         ::::::|||||:::|||:::   :::||||||:::|||   ||||||::
     583 AATATCACCAGCATTCCTCAAGGTCTTCCTCCTTCCCTTACGGAATTACA 632

     115 gValSerGlyAsnAsnLeu..............ThrSerLeuArgAlaL 127
         :::::::||||||::::::               :::||||||::::::|
     633 TCTTGATGGCAACAAAATCAGCAGAGTTGATGCAGCTAGCCTGAAAGGAC 682

     127 euProSerGlyLeuGlnLysLeu...TrpAlaTyrAsnAsnArgLeuThr 142
         ||   ::::::|||   ||||||      ::::::|||:::      :::
     683 TG...AATAATTTGGCTAAGTTGGGATTGAGTTTCAACAGCATCTCTGCT 729

     143 ............SerLeuProGluMetSerProGlyLeuGlnGluLeuAs 155
                     ||||||::::::   :::|||   |||:::||||||::
     730 GTTGACAATGGCTCTCTGGCCAAC...ACGCCTCATCTGAGGGAGCTTCA 776

     155 pValSerHisAsnGlnLeuThrArgLeuProGlnSerLeuThrGlyLeuS 172
         ::::::::::|||:::||||||||||:::|||   :::|||:::
     777 CTTGGACAACAACAAGCTTACCAGAGTACCTGGTGGGCTGGCAGAGCATA 826

     172 erSerAlaAlaArgValTyrLeuAspGlyAsnProLeuSerValArgThr 188
                     |||||||||:::::::|||   :::||||||
     827 AGTACATCCAGGTTGTCTACCTTCATAACAACAATATCTCTGTAGTTGGA 876

     189 ArgAspArgLeuCys......GlyHisHis 196
           :::   :::|||       ||||||:::
     877 TCAAGTGACTTCTGCCCACCTGGACACAAC 906

seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-619-916-1

seq_documentation_block:
; Sequence 1, Application US/08619916
; Patent No. 5763276
;  GENERAL INFORMATION:
;    APPLICANT:  Craig, William S.
;    APPLICANT:  Harper, John R.
;    APPLICANT:  Hernandez, Sam D.
;    APPLICANT:  Kostel, Paul J.
;    APPLICANT:  Parker, Jonathan R.
;    APPLICANT:  Vedvick, Thomas S.
;    TITLE OF INVENTION:  Processes for the Purification of Human
;    TITLE OF INVENTION:  Recombinant Decorin and the Detection of Guanidinium Ions
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell and Flores
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/619,916
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/272,919
;      FILING DATE:  08-JUL-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-LA 1040
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1026 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1026
US-08-619-916-1
```

```
alignment_scores:
            Quality:  139.50              Length:    260
              Ratio:  0.962                 Gaps:     14
 Percent Similarity:  55.769   Percent Identity: 25.769

alignment_block:
 US-09-068-804-14 x US-08-619-916-1   ..

 Align seg 1/1  to: US-08-619-916-1  from: 1  to: 1026

      11 ValLeuAsnValGlyAlaSerGlyLeuThrThrLeuProAspArgLeuPr 27
         |||::::::   ::::::   ||||||   :::::::|||      |||||
     142 GTGGTCCAGTGTTCTGATTTGGGTCTGGACAAAGTGCCAAAGGATCTTCC 191

      27 oProHisIleThrThrLeuValIleProAspAsnAsnLeuThrSerLeuP 44
         ||||::::::|||   |||   :::::::::|||:::::::|||
     192 CCCTGACACAACTCTGCTAGACCTGCAAAACAACAAAATAACC....... 234

      44 roGluLeuProGluGly...........LeuArgGluLeuGlu...... 54
           |||:::   :::|||           |||:::::|||:::
     235 ..GAAATCAAAGATGGAGACTTTAAGAACCTGAAGAACCTTCACGCATTG 282

      55 ......ValSerGlyAsnLeu.......................... 59
               |||::::::::::::
     283 ATTCTTGTCAACAATAAAATTAGCAAAGTTAGTCCTGGAGCATTTACACC 332

      60 ..................................GlnLeuThrSerL 64
                                          ||||||:::   |
     333 TTTGGTGAAGTTGGAACGACTTTATCTGTCCAAGAATCAGCTGAAGGAAT 382

      64 euPro...SerLeuProGlnGlyLeuGlnLysLeuTrpAlaTyrAsnAsn 79
         |||||      :::|||:::   ||||||   |||:::|||:::::|||
     383 TGCCAGAAAAAATGCCCAAAACTCTTCAGGAGCTGCGTGCCCATGAGAAT 432

      80 TrpLeuAlaSerLeuProThrLeuPro...ProGlyLeuGlyAspLeuAl 95
           ::::::   :::   :::::::::      ||||||:::::::::
     433 GAGATCACCAAAGTGCGAAAAGTTACTTTCAATGGACTGAACCAGATGAT 482

      95 aVal..................................................... 96
          |||
     483 TGTCATAGAACTGGGCACCAATCCGCTGAAGAGCTCAGGAATTGAAAATG 532

      97 ..........................................SerAsnAsn 99
                                                 :::::::::
     533 GGGCTTTCCAGGGAATGAAGAAGCTCTCCTACATCCGCATTGCTGATACC 582

     100 GlnLeuThrSerLeuProGlu...MetProProAlaLeuArgGluLeuAr 115
         ::::::|||||::::|||:::   :::||||||:::|||   ||||||::
     583 AATATCACCAGCATTCCTCAAGGTCTTCCTCCTTCCCTTACGGAATTACA 632

     115 gValSerGlyAsnAsnLeu..............ThrSerLeuArgAlaL 127
         :::::::||||||::::::               :::|||||:::::::|
     633 TCTTGATGGCAACAAAATCAGCAGAGTTGATGCAGCTAGCCTGAAAGGAC 682

     127 euProSerGlyLeuGlnLysLeu...TrpAlaTyrAsnAsnArgLeuThr 142
         ||   :::::::|||   ||||||      ::::::|||:::      :::
     683 TG...AATAATTTGGCTAAGTTGGGATTGAGTTTCAACAGCATCTCTGCT 729

     143 ............SerLeuProGluMetSerProGlyLeuGlnGluLeuAs 155
                     ||||||:::::   :::|||   |||:::||||||::
     730 GTTGACAATGGCTCTCTGGCCAAC...ACGCCTCATCTGAGGGAGCTTCA 776

     155 pValSerHisAsnGlnLeuThrArgLeuProGlnSerLeuThrGlyLeuS 172
         ::::::::::|||:::||||||||||:::|||   :::|||:::
     777 CTTGGACAACAACAAGCTTACCAGAGTACCTGGTGGGCTGGCAGAGCATA 826

     172 erSerAlaAlaArgValTyrLeuAspGlyAsnProLeuSerValArgThr 188
                     ||||||||||:::::|||   :::||||||
     827 AGTACATCCAGGTTGTCTACCTTCATAACAACAATATCTCTGTAGTTGGA 876

     189 ArgAspArgLeuCys......GlyHisHis 196
         :::   :::|||      |||||||:::
     877 TCAAGTGACTTCTGCCCACCTGGACACAAC 906
```

```
seq_name: /cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:PCT-US95-08542-1

seq_documentation_block:
; Sequence 1, Application PC/TUS9508542
;  GENERAL INFORMATION:
;    APPLICANT:  La Jolla Cancer Research Foundation
;    TITLE OF INVENTION:  Processes for the Purification of
;    TITLE OF INVENTION:  Human
;    TITLE OF INVENTION:  Recombinant Decorin and the Detection of Guanidinium
;    TITLE OF INVENTION:  Ions
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell and Flores
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/08542
;      FILING DATE:  07-JUL-1995
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Perkins, Susan M.
;      REGISTRATION NUMBER:  36,405
;      REFERENCE/DOCKET NUMBER:  FP-LA 1751
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1026 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1026
PCT-US95-08542-1


alignment_scores:
            Quality:  139.50              Length:    260
              Ratio:  0.962                 Gaps:     14
 Percent Similarity:  55.769   Percent Identity: 25.769

alignment_block:
 US-09-068-804-14 x PCT-US95-08542-1   ..

 Align seg 1/1  to: PCT-US95-08542-1  from: 1  to: 1026

      11 ValLeuAsnValGlyAlaSerGlyLeuThrThrLeuProAspArgLeuPr 27
         |||::::::   ::::::   ||||||   ::::::|||      |||||
     142 GTGGTCCAGTGTTCTGATTTGGGTCTGGACAAAGTGCCAAAGGATCTTCC 191

      27 oProHisIleThrThrLeuValIleProAspAsnAsnLeuThrSerLeuP 44
         ||||::::::|||   |||   :::::::::|||:::::::|||
     192 CCCTGACACAACTCTGCTAGACCTGCAAAACAACAAAATAACC....... 234

      44 roGluLeuProGluGly...........LeuArgGluLeuGlu...... 54
           |||:::   :::|||           |||:::::|||:::
     235 ..GAAATCAAAGATGGAGACTTTAAGAACCTGAAGAACCTTCACGCATTG 282

      55 ......ValSerGlyAsnLeu.......................... 59
               |||::::::::::::
     283 ATTCTTGTCAACAATAAAATTAGCAAAGTTAGTCCTGGAGCATTTACACC 332
```

```
      60 .....GlnLeuThrSerL 64
                        |||||||:::   |
     333 TTTGGTGAAGTTGGAACGACTTTATCTGTCCAAGAATCAGCTGAAGGAAT 382

      64 euPro...SerLeuProGlnGlyLeuGlnLysLeuTrpAlaTyrAsnAsn 79
         |||||      :::|||:::   ||||||   |||:::|||:::::::|||
     383 TGCCAGAAAAAATGCCCAAAACTCTTCAGGAGCTGCGTGCCCATGAGAAT 432

      80 TrpLeuAlaSerLeuProThrLeuPro...ProGlyLeuGlyAspLeuAl 95
            ::::::   :::   :::::::::      ||||||:::::::::
     433 GAGATCACCAAAGTGCGAAAAGTTACTTTCAATGGACTGAACCAGATGAT 482

      95 aVal..................................................... 96
          |||
     483 TGTCATAGAACTGGGCACCAATCCGCTGAAGAGCTCAGGAATTGAAAATG 532

      97 .....SerAsnAsn 99
                        :::::::::
     533 GGGCTITCCAGGGAATGAAGAAGCTCTCCTACATCCGCATTGCTGATACC 582

     100 GlnLeuThrSerLeuProGlu...MetProProAlaLeuArgGluLeuAr 115
         ::::::|||||:::|||:::   :::|||||||:::|||   ||||||::
     583 AATATCACCAGCATTCCTCAAGGTCTTCCTCCTTCCCTTACGGAATTACA 632

     115 gValSerGlyAsnAsnLeu.............ThrSerLeuArgAlaL 127
         :::::::||||||:::::              :::||||||:::::|
     633 TCTTGATGGCAACAAAATCAGCAGAGTTGATGCAGCTAGCCTGAAAGGAC 682

     127 euProSerGlyLeuGlnLysLeu...TrpAlaTyrAsnAsnArgLeuThr 142
         ||   :::::::|||   ||||||      ::::::|||:::      :::
     683 TG...AATAATTTGGCTAAGTTGGGATTGAGTTTCAACAGCATCTCTGCT 729

     143 ...........SerLeuProGluMetSerProGlyLeuGlnGluLeuAs 155
                     ||||||:::::   :::|||   |||:::||||||::
     730 GTTGACAATGGCTCTCTGGCCAAC...ACGCCTCATCTGAGGGAGCTTCA 776

     155 pValSerHisAsnGlnLeuThrArgLeuProGlnSerLeuThrGlyLeuS 172
         ::::::::::|||:::||||||||||:::|||   :::|||:::
     777 CTTGGACAACAACAAGCTTACCAGAGTACCTGGTGGGCTGGCAGAGCATA 826

     172 erSerAlaAlaArgValTyrLeuAspGlyAsnProLeuSerValArgThr 188
                     ||||||||||:::::|||   :::||||||
     827 AGTACATCCAGGTTGTCTACCTTCATAACAACAATATCTCTGTAGTTGGA 876

     189 ArgAspArgLeuCys......GlyHisHis 196
             :::   :::|||       |||||||:::
     877 TCAAGTGACTTCTGCCCACCTGGACACAAC 906

seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-442-063A-44

seq_documentation_block:
; Sequence 44, Application US/08442063A
; Patent No. 5705609
;  GENERAL INFORMATION:
;    APPLICANT:  RUOSLAHTI, ERKKI I.
;    APPLICANT:  PIERSCHBACHER, MICHAEL D.
;    APPLICANT:  CARDENAS, JOSE
;    APPLICANT:  CRAIG, WILLIAM
;    APPLICANT:  MULLEN, DANIEL G.
;    TITLE OF INVENTION:  DECORIN FRAGMENTS AND METHODS OF
;    TITLE OF INVENTION:  INHIBITING CELL REGULATORY FACTORS
;    NUMBER OF SEQUENCES:  62
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CAMPBELL AND FLORES
;      STREET:  4370 LA JOLLA VILLAGE DRIVE, SUITE 700
;      CITY:  SAN DIEGO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/442,063A
;      FILING DATE:  16-MAY-1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/865,652
;      FILING DATE:  03-APR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CAMPBELL, CATHRYN A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-LA 1454
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-535-9001
;      TELEFAX:  619-535-8949
;  INFORMATION FOR SEQ ID NO:  44:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  849 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(1..840, 844..849)
US-08-442-063A-44


alignment_scores:
            Quality:  135.50              Length:    248
              Ratio:  0.975                 Gaps:     13
 Percent Similarity:  56.048   Percent Identity: 25.806

alignment_block:
 US-09-068-804-14 x US-08-442-063A-44   ..

 Align seg 1/1  to: US-08-442-063A-44  from: 1  to: 849

      11 ValLeuAsnValGlyAlaSerGlyLeuThrThrLeuProAspArgLeuPr 27
         |||::::::   ::::::   ||||||   ::::::|||      |||||
     106 GTGGTCCAGTGTTCTGATTTGGGTCTGGACAAAGTGCCAAAGGATCTTCC 155

      27 oProHisIleThrThrLeuValIleProAspAsnAsnLeuThrSerLeuP 44
         ||||::::::|||   |||   :::::::::|||:::::::|||
     156 CCCTGACACAACTCTGCTAGACCTGCAAAACAACAAAATAACC....... 198

      44 roGluLeuProGluGly.....LeuArgGluLeuGlu..... 54
         |||:::  ::::||        |||::::::|||:::
     199 ..GAAATCAAAGATGGAGACTTTAAGAACCTGAAGAACCTTCACGCATTG 246

      55 ......ValSerGlyAsnLeu..... 59
            |||:::::::::::::
     247 ATTCTTGTCAACAATAAAATTAGCAAAGTTAGTCCTGGAGCATTTACACC 296

      60 .....GlnLeuThrSerL 64
                        |||||||:::   |
     297 TTTGGTGAAGTTGGAACGACTTTATCTGTCCAAGAATCAGCTGAAGGAAT 346

      64 euPro...SerLeuProGlnGlyLeuGlnLysLeuTrpAlaTyrAsnAsn 79
         |||||      :::|||:::   ||||||   |||:::|||:::::::|||
     347 TGCCAGAAAAAATGCCCAAAACTCTTCAGGAGCTGCGTGCCCATGAGAAT 396

      80 TrpLeuAlaSerLeuProThrLeuPro...ProGlyLeuGlyAspLeuAl 95
            ::::::   :::   :::::::::      ||||||:::::::::
     397 GAGATCACCAAAGTGCGAAAAGTTACTTTCAATGGACTGAACCAGATGAT 446

      95 aVal..................................................... 96
          |||
     447 TGTCATAGAACTGGGCACCAATCCGCTGAAGAGCTCAGGAATTGAAAATG 496

      97 .....SerAsnAsn 99
                        :::::::::
     497 GGGCTTTCCAGGGAATGAAGAAGCTCTCCTACATCCGCATTGCTGATACC 546
```

```
    100 GlnLeuThrSerLeuProGlu...MetProProAlaLeuArgGluLeuAr 115
        ::::::||||||:::|||:::   :::|||||||:::|||   ||||||::
    547 AATATCACCAGCATTCCTCAAGGICTTCCTCCTTCCCTTACGGAATTACA 596

    115 gValSerGlyAsnAsnLeu..............ThrSerLeuArgAlaL 127
        :::::::||||||:::::::               :::|||||||:::::::|
    597 TCTTGATGGCAACAAAATCAGCAGAGTTGATGCAGCTAGCCTGAAAGGAC 646

    127 euProSerGlyLeuGlnLysLeu...TrpAlaTyrAsnAsnArgLeuThr 142
        ||   :::::|||   ||||||      :::::|||:::      :::
    647 TG...AATAATTTGGCTAAGTTGGGATTGAGTTTCAACAGCATCTCTGCT 693

    143 ............SerLeuProGluMetSerProGlyLeuGlnGluLeuAs 155
                    ||||||::::::   :::|||   |||:::||||||::
    694 GTTGACAATGGCTCTCTGGCCAAC...ACGCCTCATCTGAGGGAGCTTCA 740

    155 pValSerHisAsnGlnLeuThrArgLeuProGlnSerLeuThrGlyLeuS 172
        ::::::::::||:::|||||||||||:::|||   :::|||:::
    741 CTTGGACAACAACAAGCTTACCAGAGTACCTGGTGGGCTGGCAGAGCATA 790

    172 erSerAlaAlaArgValTyrLeuAspGlyAsnProLeuSerVal 186
                    |||||||||:::::|||   :::||||||
    791 AGTACATCCAGGTTGTCTACCTTCATAACAACAATATCTCTGTA 834

seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-442-063A-41

seq_documentation_block:
; Sequence 41, Application US/08442063A
; Patent No. 5705609
;  GENERAL INFORMATION:
;    APPLICANT:  RUOSLAHTI, ERKKI I.
;    APPLICANT:  PIERSCHBACHER, MICHAEL D.
;    APPLICANT:  CARDENAS, JOSE
;    APPLICANT:  CRAIG, WILLIAM
;    APPLICANT:  MULLEN, DANIEL G.
;    TITLE OF INVENTION:  DECORIN FRAGMENTS AND METHODS OF
;    TITLE OF INVENTION:  INHIBITING CELL REGULATORY FACTORS
;    NUMBER OF SEQUENCES:  62
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CAMPBELL AND FLORES
;      STREET:  4370 LA JOLLA VILLAGE DRIVE, SUITE 700
;      CITY:  SAN DIEGO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/442,063A
;      FILING DATE:  16-MAY-1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/865,652
;      FILING DATE:  03-APR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CAMPBELL, CATHRYN A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-LA 1454
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-535-9001
;      TELEFAX:  619-535-8949
;  INFORMATION FOR SEQ ID NO:  41:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  711 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(1..702, 706..711)
US-08-442-063A-41


alignment_scores:
            Quality:  127.00              Length:    216
              Ratio:  1.050                 Gaps:      9
 Percent Similarity:  56.019   Percent Identity: 25.463

alignment_block:
 US-09-068-804-14 x US-08-442-063A-41   ..

 Align seg 1/1  to: US-08-442-063A-41  from: 1  to: 711

     11 ValLeuAsnValGlyAlaSerGlyLeuThrThrLeuProAspArgLeuPr 27
        |||::::::   ::::::   ||||||   ::::::|||      |||||
    106 GTGGTCCAGTGTTCTGATTTGGGTCTGGACAAAGTGCCAAAGGATCTTCC 155

     27 oProHisIleThrThrLeuValIleProAspAsnAsnLeuThrSerLeuP 44
        ||||::::::|||   |||   :::::::::|||:::::::|||   :::
    156 CCCTGACACAACTCTGCTAGACCTGCAAAACAACAAAATAACCGAAATCA 205

     44 roGluLeuProGluGlyLeuArgGluLeuGluValSerGlyAsnLeuGln 60
          :::   :::::::::::|||            :::|||
    206 AAGAT......GGAGACTTTAAGAACCTGAAGAACCTTCACGCATTGATT 249

     61 LeuThrSerLeuProSerLeuProGlnGlyLeuGlnLysLeuTrpAlaTy 77
        |||::::::
    250 CTTGTCAAC..................................... 258

     77 rAsnAsnTrpLeuAlaSerLeuProThrLeuProProGly..... 90
                  :::   ::::::::::::::::||||||
    259 .............AATAAAATTAGCAAAGTTAGTCCTGGAGCATTTACAC 295

     91 ...........LeuGlyAspLeuAlaValSerAsnAsnGlnLeuThrSer 103
                   |||      |||   :::|||:::||||||||||:::
    296 CTTTGGTGAAGTTGGAACGACTTTATCTGTCCAAGAATCAGCTGAAGGAA 345

    104 LeuProGlu...MetProProAlaLeuArgGluLeuArgValSerGlyAs 119
        |||||||||   ||||||   :::|||:::||||||||||:::      ||
    346 TTGCCAGAAAAAATGCCCAAAACTCTTCAGGAGCTGCGTGCCCATGAGAA 395

    119 nAsnLeuThrSerLeuArgAlaLeuPro..................... 128
        |::::::|||   :::|||   ::::::
    396 TGAGATCACCAAAGTGCGAAAAGTTACTTTCAATGGACTGAACCAGATGA 445

    128 ............................................... 128

    446 TTGTCATAGAACTGGGCACCAATCCGCTGAAGAGCTCAGGAATTGAAAAT 495

    129 .........SerGlyLeuGlnLysLeuTrpAlaTyr............As 138
                  |||::::::|||||   :::|||            ::
    496 GGGGCTTTCCAGGGAATGAAGAAGCTC...TCCTACATCCGCATTGCTGA 542

    138 nAsnArgLeuThrSerLeuProGlu...MetSerProGlyLeuGlnGluL 154
        ::::   :::||||||:::|||:::   :::::::|||:::|||   ||||
    543 TACCAATATCACCAGCATTCCTCAAGGTCTTCCTCCTTCCCTTACGGAAT 592

    154 euAspValSerHisAsnGlnLeuThrArgLeu...ProGlnSerLeuThr 169
        ||:::::::::   |||:::::::::|||:::   :::   ||||||:::
    593 TACATCTTGATGGCAACAAAATCAGCAGAGTTGATGCAGCTAGCCTGAAA 642

    170 GlyLeuSerSerAlaAlaArgValTyrLeuAspGlyAsnProLeuSer 185
        ||||||::::::   |||:::::   |||:::   |||:::::::|||
    643 GGACTGAATAATTTGGCTAAGTTGGGATTGAGTTTCAACAGCATCTCT 690

seq_name: /cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:PCT-US94-07644A-1

seq_documentation_block:
; Sequence 1, Application PC/TUS9407644A
;  GENERAL INFORMATION:
;    APPLICANT:  COR Therapeutics, Inc.
```

```
;    TITLE OF INVENTION:  Platelet Glycoprotein V Gene and Uses
;    NUMBER OF SEQUENCES:  43
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend Khourie and Crew
;      STREET:  379 Lytton Avenue
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  US
;      ZIP:  94301
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/07644A
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dow, Karen B.
;      REGISTRATION NUMBER:  29,684
;      REFERENCE/DOCKET NUMBER:  012418-003000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 326-2400
;      TELEFAX:  (415) 326-2422
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7452 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    FEATURE:
;      NAME/KEY:  intron
;      LOCATION:  1462..2419
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2422..4101
;    FEATURE:
;      NAME/KEY:  misc_signal
;      LOCATION:  68..76
;      OTHER INFORMATION:  /function= "Putitive TPA responsive
;      OTHER INFORMATION:  element"
;      OTHER INFORMATION:  /label= TRE
;    FEATURE:
;      NAME/KEY:  misc_signal
;      LOCATION:  471..478
;      OTHER INFORMATION:  /function= "Ets-1 cis-acting
;      OTHER INFORMATION:  sequence"
;      OTHER INFORMATION:  /label= Ets-1
;    FEATURE:
;      NAME/KEY:  misc_signal
;      LOCATION:  493..502
;      OTHER INFORMATION:  /function= "Ets-1 cis-acting
;      OTHER INFORMATION:  sequence"
;      OTHER INFORMATION:  /label= Ets-1
;    FEATURE:
;      NAME/KEY:  repeat_region
;      LOCATION:  593..881
;      OTHER INFORMATION:  /rpt_type= "other"
;      OTHER INFORMATION:  /label= Alu
;    FEATURE:
;      NAME/KEY:  misc_signal
;      LOCATION:  897..904
;      OTHER INFORMATION:  /function= "Ets-1 cis-acting
;      OTHER INFORMATION:  sequence"
;      OTHER INFORMATION:  /label= Ets-1
;    FEATURE:
;      NAME/KEY:  misc_binding
;      LOCATION:  1142..1149
;      OTHER INFORMATION:  /function= "Sp1 binding site"
;      OTHER INFORMATION:  /standard_name= "Sp1"
;    FEATURE:
;      NAME/KEY:  misc_signal
;      LOCATION:  1178..1184
;      OTHER INFORMATION:  /function= "Ets-1 cis-acting
;      OTHER INFORMATION:  sequences"
;      OTHER INFORMATION:  /label= Ets-1
;    FEATURE:
;      NAME/KEY:  TATA_signal
;      LOCATION:  1199..1203
;    FEATURE:
;      NAME/KEY:  TATA_signal
;      LOCATION:  1263..1269
;    FEATURE:
;      NAME/KEY:  misc_binding
;      LOCATION:  1285..1289
;      OTHER INFORMATION:  /function= "GATA-1 binding site"
;    FEATURE:
;      NAME/KEY:  misc_binding
;      LOCATION:  1321..1326
;      OTHER INFORMATION:  /function= "GATA-1 binding site"
;    FEATURE:
;      NAME/KEY:  misc_signal
;      LOCATION:  1365..1372
;      OTHER INFORMATION:  /function= "Ets-1 cis-acting
;      OTHER INFORMATION:  sequences"
;      OTHER INFORMATION:  /label= Ets-1
;    FEATURE:
;      NAME/KEY:  repeat_region
;      LOCATION:  6133..6440
;      OTHER INFORMATION:  /rpt_type= "other"
;      OTHER INFORMATION:  /label= Alu
;    FEATURE:
;      NAME/KEY:  misc_signal
;      LOCATION:  5610..5615
;      OTHER INFORMATION:  /standard_name= "Polyadenylation
;      OTHER INFORMATION:  signal sequence"
;    FEATURE:
;      NAME/KEY:  misc_signal
;      LOCATION:  6966..6971
;      OTHER INFORMATION:  /standard_name= "Polyadenylation
;      OTHER INFORMATION:  signal sequence"
;    FEATURE:
;      NAME/KEY:  misc_signal
;      LOCATION:  7224..7229
;      OTHER INFORMATION:  /standard_name= "Polyadenylation
;      OTHER INFORMATION:  signal sequence"
;    FEATURE:
;      NAME/KEY:  misc_signal
;      LOCATION:  7358..7363
;      OTHER INFORMATION:  /standard_name= "Polyadenylation
;      OTHER INFORMATION:  signal sequence"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..7452
;      OTHER INFORMATION:  /standard_name= "Nucleotide
;      OTHER INFORMATION:  sequence containing the human GPV gene"
PCT-US94-07644A-1


alignment_scores:
            Quality:  123.50              Length:    235
              Ratio:  0.957                 Gaps:     14
 Percent Similarity:  54.894   Percent Identity: 28.936

alignment_block:
 US-09-068-804-14 x PCT-US94-07644A-1   ..

 Align seg 1/1  to: PCT-US94-07644A-1  from: 1  to: 7452

      11 ValLeuAsnValGlyAlaSerGlyLeuThrThrLeuPro..... 23
         :::|||:::::::::::::::::::|||||   ||||||
    2941 TTGTTGGATTTATCGGGAAACAACCTGACCCACCTGCCCAAGGGGTTGCT 2990
```

```
     24 ..................AspArgLeuProProHisIleThrThrLeuV 34
                          :::||||||      |||   :::   ||||
   2991 TGGAGCACAGGCTAAGCTCGAGAGACTTCTGCTCCACTCGAACCGCCTTG 3040

     34 alIleProAsp............AsnAsnLeuThrSerLeuProGluLeu 46
        ||      |||             |||:::|||   :::|||:::||||||
   3041 TGTCTCTGGATTCGGGGCTGTTGAACAGCCTGGGCGCCCTGACGGAGCTG 3090

     47 Pro.........GluGlyLeuArgGluLeuGluValSerGlyAsnLeuGl 60
        :::          :::   :::|||   ::::::   ::::::      ::
   3091 CAGTTCCACCGAAATCACATCCGTTCCATCGCACCCGGGGCCTTCGACCG 3140

     60 nLeuThrSerLeuProSerLeuProGlnGlyLeuGlnLysLeuTrpAlaT 77
        :|||:::::|||:::||||||:::   :::
   3141 GCTCCCAAACCTCAGTTCTTTGACGCTTTCG..................... 3171

     77 yrAsnAsnTrpLeuAlaSerLeuProThr............LeuProPro 89
           |||   ||||||   ||||||::                 :::
   3172 ..AGAAACCACCTTGCGTTTCTCCCCTCTGCGCTCTTTCTTCATTCGCAC 3219

     90 GlyLeuGlyAspLeuAlaValSerAsnAsnGlnLeuThrSerLeuPro.. 105
        :::|||      |||::::::   :::|||:::|||:::   ||||||
   3220 AATCTGACTCTGTTGACTCTGTTCGAGAACCCGCTGGCAGAGCTCCCGGG 3269

    106 .............GluMetProProAlaLeuArgGluLeuArgValSerG 118
                      ||||||      :::|||:::||||||::::::
   3270 GGTGCTCTTCGGGGAGATG...GGGGGCCTGCAGGAGCTGTGGCTG.... 3312

    118 lyAsnAsnLeuThrSerLeuArgAlaLeuProSerGly............. 130
           |||   |||   ||||||:::|||||||:::::
   3313 .....AACCGCACCCAGCTGCGCACCCTGCCCGCCGCCGCCTTCCGCAAC 3357

    131 LeuGlnLysLeuTrp............AlaTyrAsnAsnArgLeuThrSe 143
        |||   :::|||:::             :::  :::   ||||||:::::
   3358 CTGAGCCGCCTGCGGTACTTAGGGGTGACTCTGAGCCCGCGGCTGAGCGC 3407

    143 rLeuProGluMetSer............ProGlyLeuGlnGluLeuAspV 156
        :||||||:::   :::                  ||||||   |||:::
   3408 GCTTCCGCAGGGCGCCTTCCAGGGCCTTGGCGAGCTCCAGGTGCTCGCCC 3457

    156 alSerHisAsnGlnLeuThrArgLeuProGlnSer...LeuThrGlyLeu 171
        ::      |||   ||||||   ||||||::::::   |||   ||||||
   3458 TGCACTCCAACGGCCTGACCGCCCTCCCCGACGGCTTGCTGCGCGGCCTC 3507

    172 SerSerAlaAlaArgValTyrLeuAspGlyAsnProLeuSerValArgTh 188
        :::          :::|||   |||
   3508 GGCAAGCTGCGCCAGGTGTCCCTG.......................CG 3533

    188 rArgAspArgLeuCysGlyHisHisTrpPro...PheArgHisGlnAspT 204
         |||:::||||||   :::      ::::::   ||||||:::   ::::
   3534 CCGCAACAGGCTGCGCGCCCTGCCCCGTGCCCTCTTCCGCAATCTCAGCA 3583

    204 hrLeu 205
        ::|||
   3584 GCCTG 3588

seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-365-486A-6

seq_documentation_block:
; Sequence 6, Application US/08365486A
; Patent No. 5834306
;  GENERAL INFORMATION:
;    APPLICANT:  Webster, Keith A.
;    APPLICANT:  Bishopric, Nanette H.
;    TITLE OF INVENTION:  Tissue Specific Hypoxia Regulated
;    TITLE OF INVENTION:  Therapeutic Constructs
;    NUMBER OF SEQUENCES:  31
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dehlinger & Associates
;      STREET:  350 Cambridge Avenue, Suite 250
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94306
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/365,486A
;      FILING DATE:  23-DEC-1994
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sholtz, Charles K.
;      REGISTRATION NUMBER:  38,615
;      REFERENCE/DOCKET NUMBER:  8255-0018
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 324-0880
;      TELEFAX:  (415) 324-0960
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  240 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      INDIVIDUAL ISOLATE:  HREE1 (Hypoxia responsive enhancer
;      INDIVIDUAL ISOLATE:  element 1)
US-08-365-486A-6


alignment_scores:
            Quality:   95.00              Length:     76
              Ratio:  1.759                Gaps:      0
 Percent Similarity:  71.053   Percent Identity: 27.632

alignment_block:
 US-09-068-804-14 x US-08-365-486A-6   ..

 Align seg 1/1  to: US-08-365-486A-6  from: 1  to: 240

     74 LeuTrpAlaTyrAsnAsnTrpLeuAlaSerLeuProThrLeuProProGl 90
        |||:::|||   ::::::   |||:::::|||:::|||      |||||
      7 CTACGTGCTGTCTCACACAGCCTGTCTGACCTCTCGACCTACCGGCCCGG 56

     90 yLeuGlyAspLeuAlaValSerAsnAsnGlnLeuThrSerLeuProGluM 107
        |:::   :::|||   :::   ::::::   |||::::::|||:::
     57 GATCCCGGCCCTACGTGCTGTCTCACACAGCCTGTCTGACCTCTCGACCT 106

    107 etProProAlaLeuArgGluLeuArgValSerGlyAsnAsnLeuThrSer 123
            |||::::::   :::|||||||:::   :::::::::|||::::::
    107 ACCGGCCCGGGATCCCGGCCCTACGTGCTGTCTCACACAGCCTGTCTGAC 156

    124 LeuArgAlaLeuProSerGlyLeuGlnLysLeuTrpAlaTyrAsnAsnAr 140
        |||   :::      :::|||::::::   |||:::|||   ::::::
    157 CTCTCGACCTACCGGCCCGGGATCCCGGCCCTACGTGCTGTCTCACACAG 206

    140 gLeuThrSerLeuProGluMetSerPro 149
         |||::::::|||:::         |||
    207 CCTGTCTGACCTCTCGACCTACCGGCCG 234

seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-452-531-2

seq_documentation_block:
; Sequence 2, Application US/08452531
; Patent No. 5733778
;  GENERAL INFORMATION:
;    APPLICANT:  MATLASHEWSKI, Gregory
;    APPLICANT:  CHAREST, Hugues
;    TITLE OF INVENTION:  GENES OF LEISHMANIA WHICH ARE DIFFERENTIALLY-
;    TITLE OF INVENTION:  EXPRESSED IN AMASTIGOTE FORM
;    NUMBER OF SEQUENCES:  4
```

```
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sim & McBurney
;      STREET:  Suite 701, 330 University Avenue
;      CITY:  Toronto
;      STATE:  Ontario
;      COUNTRY:  Canada
;      ZIP:  M5G 1R7
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/452,531
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/115,987
;      FILING DATE:  03-SEP-1993
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  STEWART, Michael I.
;      REGISTRATION NUMBER:  24,973
;      REFERENCE/DOCKET NUMBER:  1038-448 MIS:bh
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (416) 595-1155
;      TELEFAX:  (416) 595-1163
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  711 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-452-531-2


alignment_scores:
            Quality:  100.50              Length:    182
              Ratio:  0.897                 Gaps:      4
 Percent Similarity:  61.538   Percent Identity: 23.077

alignment_block:
 US-09-068-804-14 x US-08-452-531-2   ..

 Align seg 1/1  to: US-08-452-531-2  from: 1  to: 711

      12 LeuAsnValGlyAlaSerGlyLeuThrThrLeuProAspArgLeuProPr 28
         |||:::||||||:::   ::::::   :::|||:::          |||:: 
     112 CTCTCCGTTGGCCCGCAGTCCGTCGGCCCGCTCTCTGTTGGC...CCGCA 158

      28 oHisIleThrThrLeuValIleProAspAsnAsnLeuThrSerLeuProG 45
         :   :::   :::|||   :::       :::::::::   :::|||:::  
     159 GGCTGTTGGCCCGCTCTCCGTTGGCCCGCAGTCCGTCGGCCCGCTCTCTG 208

      45 luLeuProGluGlyLeuArgGluLeuGluValSer.....Gly 57
            |||::::::::::      |||   |||:::          |||  
     209 TTGGCCCGCAGGCTGTTGGCCCGCTCTCTGTTGGCCCGCAGTCCGTTGGC 258

      58 AsnLeuGlnLeuThrSerLeuProSerLeuProGlnGlyLeuGlnLysLe 74
          |||   :::   :::|||:::      ||||||::::::      ||  
     259 CCGCTCTCCGTTGGCCCGCTCTCCGTTGGCCCGCAGTCTGTTGGCCCGCT 308

      74 uTrpAlaTyrAsnAsnTrpLeuAlaSerLeuProThrLeuProProGlyL 91
         |   :::   ::::::   ::::::::::|||:::::::   |||:::::::  
     309 CTCCGTTGGCTCGCAGTCCGTCGGCCCGCTCTCTGTTGGTCCGCAGTCCG 358

      91 euGlyAspLeuAlaValSerAsnAsnGlnLeuThrSerLeuProGluMet 107
         ::|||   |||:::|||:::   :::   :::   :::|||:::  
     359 TCGGCCCGCTCTCCGTTGGCCCGCAGGCTGTTGGCCCGCTCTCCGTTGGC 408

     108 ProProAlaLeuArgGluLeuArgValSerGlyAsnAsnLeuThrSerLe 124
         |||:::::::::      |||   |||:::   :::::::::   :::||  
     409 CCGCAGTCCGTCGGCCCGCTCTCTGTTGGCCCGCAGGCTGTTGGCCCGCT 458

     124 uArgAlaLeuProSerGlyLeuGlnLysLeuTrpAlaTyrAsnAsnArgL 141
         |   :::   |||   ::::::      |||   :::      :::   :  
     459 CTCTGTTGGCCCGCAGTCCGTTGGCCCGCTCTCCGTTGGCCCGCAGTCTG 508

     141 euThrSerLeuProGluMetSerProGlyLeuGlnGluLeuAspValSer 157
         ::   :::|||:::      |||:::::::::      |||:::|||:::  
     509 TTGGCCCGCTCTCCGTTGGCTCGCAGTCCGTCGGCCCGCTCTCTGTTGGT 558

     158 HisAsnGlnLeuThrArgLeu.........ProGlnSerLeuThrGlyLe 171
            :::   :::      |||         |||||||||:::      ||  
     559 CCGCAGTCCGTCGGCCCGCTCTCCGTTGGCCCGCAGTCTGTCGGCCCGCT 608

     171 uSerSerAlaAlaArgValTyrLeuAspGlyAsnProLeuSerVal 186
         ||||   :::::::::         :::   :::||||||||||||  
     609 CTCCGTTGGCCCGCAG.........TCCGTCGGCCCGCTCTCCGTT 645

seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-460-746A-2

seq_documentation_block:
; Sequence 2, Application US/08460746A
; Patent No. 5780591
;  GENERAL INFORMATION:
;    APPLICANT:  MATLASHEWSKI, Gregory
;    APPLICANT:  CHAREST, Hugues
;    TITLE OF INVENTION:  DIFFERENTIALLY EXPRESSED LEISHMANIA
;    TITLE OF INVENTION:  GENES AND PROTEINS
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sim & McBurney
;      STREET:  6th Floor, 330 University Avenue
;      CITY:  Toronto
;      STATE:  Ontario
;      COUNTRY:  Canada
;      ZIP:  M5G 1R7
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/460,746A
;      FILING DATE:  02-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/302,463
;      FILING DATE:  12-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/115,987
;      FILING DATE:  03-SEP-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  STEWART, Michael I.
;      REGISTRATION NUMBER:  24,973
;      REFERENCE/DOCKET NUMBER:  1038-483 MIS:jb
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (416) 595-1155
;      TELEFAX:  (416) 595-1163
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  711 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-460-746A-2


alignment_scores:
            Quality:  100.50              Length:    182
              Ratio:  0.897                 Gaps:      4
 Percent Similarity:  61.538   Percent Identity: 23.077

alignment_block:
 US-09-068-804-14 x US-08-460-746A-2   ..
```

```
Align seg 1/1  to: US-08-460-746A-2  from: 1  to: 711

     12 LeuAsnValGlyAlaSerGlyLeuThrThrLeuProAspArgLeuProPr 28
     112 CTCTCCGTTGGCCCGCAGTCCGTCGGCCCGCTCTCTGTTGGC...CCGCA 158

      28 oHisIleThrThrLeuValIleProAspAsnAsnLeuThrSerLeuProG 45
     159 GGCTGTTGGCCCGCTCTCCGTTGGCCCGCAGTCCGTCGGCCCGCTCTCTG 208

      45 luLeuProGluGlyLeuArgGluLeuGluValSer.....Gly 57
     209 TTGGCCCGCAGGCTGTTGGCCCGCTCTCTGTTGGCCCGCAGTCCGTTGGC 258

      58 AsnLeuGlnLeuThrSerLeuProSerLeuProGlnGlyLeuGlnLysLe 74
     259 CCGCTCTCCGTTGGCCCGCTCTCCGTTGGCCCGCAGTCTGTTGGCCCGCT 308

      74 uTrpAlaTyrAsnAsnTrpLeuAlaSerLeuProThrLeuProProGlyL 91
     309 CTCCGTTGGCTCGCAGTCCGTCGGCCCGCTCTCTGTTGGTCCGCAGTCCG 358

      91 euGlyAspLeuAlaValSerAsnAsnGlnLeuThrSerLeuProGluMet 107
     359 TCGGCCCGCTCTCCGTTGGCCCGCAGGCTGTTGGCCCGCTCTCCGTTGGC 408

     108 ProProAlaLeuArgGluLeuArgValSerGlyAsnAsnLeuThrSerLe 124
     409 CCGCAGTCCGTCGGCCCGCTCTCTGTTGGCCCGCAGGCTGTTGGCCCGCT 458

     124 uArgAlaLeuProSerGlyLeuGlnLysLeuTrpAlaTyrAsnAsnArgL 141
     459 CTCTGTTGGCCCGCAGTCCGTTGGCCCGCTCTCCGTTGGCCCGCAGTCTG 508

     141 euThrSerLeuProGluMetSerProGlyLeuGlnGluLeuAspValSer 157
     509 TTGGCCCGCTCTCCGTTGGCTCGCAGTCCGTCGGCCCGCTCTCTGTTGGT 558

     158 HisAsnGlnLeuThrArgLeu.....ProGlnSerLeuThrGlyLe 171
     559 CCGCAGTCCGTCGGCCCGCTCTCCGTTGGCCCGCAGTCTGTCGGCCCGCT 608

     171 uSerSerAlaAlaArgValTyrLeuAspGlyAsnProLeuSerVal 186
     609 CTCCGTTGGCCCCGCAG.....TCCGTCGGCCCGCTCTCCGTT 645

seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-460-555-2

seq_documentation_block:
; Sequence 2, Application US/08460555
; Patent No. 5827671
; Patent No. 5827671 5827671
;  GENERAL INFORMATION:
;    APPLICANT:  MATLASHEWSKI, Gregory
;    APPLICANT:  CHAREST, Hugues
;    TITLE OF INVENTION:  DIFFERENTIALLY EXPRESSED LEISHMANIA
;    TITLE OF INVENTION:  GENES AND PROTEINS
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sim & McBurney
;      STREET:  Suite 701, 330 University Avenue
;      CITY:  Toronto
;      STATE:  Ontario
;      COUNTRY:  Canada
;      ZIP:  M5G 1R7
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/460,555
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/302,463
;      FILING DATE:  12-SEP-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/115,987
;      FILING DATE:  03-SEP-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  STEWART, Michael I.
;      REGISTRATION NUMBER:  24,973
;      REFERENCE/DOCKET NUMBER:  1038-485 MIS:vg
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (416) 595-1155
;      TELEFAX:  (416) 595-1163
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  711 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-460-555-2


alignment_scores:
          Quality:  100.50              Length:    182
            Ratio:  0.897                 Gaps:      4
 Percent Similarity:  61.538   Percent Identity: 23.077

alignment_block:
 US-09-068-804-14 x US-08-460-555-2   ..

 Align seg 1/1  to: US-08-460-555-2  from: 1  to: 711

      12 LeuAsnValGlyAlaSerGlyLeuThrThrLeuProAspArgLeuProPr 28
     112 CTCTCCGTTGGCCCGCAGTCCGTCGGCCCGCTCTCTGTTGGC...CCGCA 158

      28 oHisIleThrThrLeuValIleProAspAsnAsnLeuThrSerLeuProG 45
     159 GGCTGTTGGCCCGCTCTCCGTTGGCCCGCAGTCCGTCGGCCCGCTCTCTG 208

      45 luLeuProGluGlyLeuArgGluLeuGluValSer.....Gly 57
     209 TTGGCCCGCAGGCTGTTGGCCCGCTCTCTGTTGGCCCGCAGTCCGTTGGC 258

      58 AsnLeuGlnLeuThrSerLeuProSerLeuProGlnGlyLeuGlnLysLe 74
     259 CCGCTCTCCGTTGGCCCGCTCTCCGTTGGCCCGCAGTCTGTTGGCCCGCT 308

      74 uTrpAlaTyrAsnAsnTrpLeuAlaSerLeuProThrLeuProProGlyL 91
     309 CTCCGTTGGCTCGCAGTCCGTCGGCCCGCTCTCTGTTGGTCCGCAGTCCG 358

      91 euGlyAspLeuAlaValSerAsnAsnGlnLeuThrSerLeuProGluMet 107
     359 TCGGCCCGCTCTCCGTTGGCCCGCAGGCTGTTGGCCCGCTCTCCGTTGGC 408

     108 ProProAlaLeuArgGluLeuArgValSerGlyAsnAsnLeuThrSerLe 124
     409 CCGCAGTCCGTCGGCCCGCTCTCTGTTGGCCCGCAGGCTGTTGGCCCGCT 458

     124 uArgAlaLeuProSerGlyLeuGlnLysLeuTrpAlaTyrAsnAsnArgL 141
     459 CTCTGTTGGCCCGCAGTCCGTTGGCCCGCTCTCCGTTGGCCCGCAGTCTG 508

     141 euThrSerLeuProGluMetSerProGlyLeuGlnGluLeuAspValSer 157
     509 TTGGCCCGCTCTCCGTTGGCTCGCAGTCCGTCGGCCCGCTCTCTGTTGGT 558

     158 HisAsnGlnLeuThrArgLeu.....ProGlnSerLeuThrGlyLe 171
```

```
          :::   :::      |||         |||||||||:::      ||
     559 CCGCAGTCCGTCGGCCCGCTCTCCGTTGGCCCGCAGTCTGTCGGCCCGCT 608

     171 uSerSerAlaAlaArgValTyrLeuAspGlyAsnProLeuSerVal 186
         ||||   :::::::::         :::   :::|||||||||||
     609 CTCCGTTGGCCCGCAG.........TCCGTCGGCCCGCTCTCCGTT 645

seq_name: /cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-666-271-4

seq_documentation_block:
; Sequence 4, Application US/08666271
; Patent No. 5920000
;  GENERAL INFORMATION:
;    APPLICANT:  JONES, JONATHAN D
;    APPLICANT:  HAMMOND-KOSACK, KIM E
;    APPLICANT:  THOMAS, COLWYN M
;    APPLICANT:  JONES, DAVID A
;    TITLE OF INVENTION:  PLANT PATHOGEN RESISTANCE GENES AND USES
;    TITLE OF INVENTION:  THEREOF
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NIXON & VANDERHYE P.C.
;      STREET:  1100 NORTH GLEBE ROAD, 8TH FLOOR
;      CITY:  ARLINGTON
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22201
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/666,271
;      FILING DATE:  19-SEP-1996
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB94/02812
;      FILING DATE:  23-DEC-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9326428.1
;      FILING DATE:  24-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9409363.0
;      FILING DATE:  11-MAY-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SADOFF, B.J.
;      REGISTRATION NUMBER:  36,663
;      REFERENCE/DOCKET NUMBER:  620-7
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-816-4091
;      TELEFAX:  703-816-4100
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2880 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-666-271-4


alignment_scores:
            Quality:  107.00              Length:    595
              Ratio:  0.375                Gaps:     29
 Percent Similarity:  47.899   Percent Identity: 18.655

alignment_block:
 US-09-068-804-14 x US-08-666-271-4   ..

 Align seg 1/1  to: US-08-666-271-4  from: 1  to: 2880

      20 ThrThrLeuProAspArgLeuProProHisIleThrThrLeuValIlePr 36

    :::|||:::|||:::   :::::::::|||:::|||||||||   :::::
676 TCCACTATTCCTTCAAATTTCTCTTCTCATTTAACAACTCTACAACTTTC 725

 36 oAspAsnAsnLeuThrSerLeuProGluLeuProGluGlyLeuArgGluL 53
    ::::::::::|||   ::::::      |||||||||   :::   :::|
726 AGGCACAGAGTTACATGGGATA......TTGCCCGAAAGAGTTTTTCACC 769

 53 euGluValSerGlyAsnLeuGlnLeuThrSerLeuProSerLeuProGln 69
    ||            :::|||:::|||:::      |||   |||:::
770 TTTCCAACTTACAATCCCTTCATTTATCAGTCAATCCCCAGCTCACGGTT 819

 70 GlyLeuGlnLysLeuTrpAlaTyrAsnAsnTrpLeuAlaSerLeuProTh 86
      :::::::::           :::::::|||   :::|||   :::
820 AGGTTTCCCACA...........ACCAAATGGAATAGCAGTGCATCACT 857

 86 rLeuProProGlyLeuGlyAspLeuAlaValSerAsn............. 98
     ::::::      ::::::::::::::::::::::::
858 CATGACGTTATACGTCGATAGTGTGAATATTGCTGATAGGATACCTAAAT 907

 99 .....AsnGlnLeuThrSerLeuProGluMet.................. 107
         ::::::|||||||||||||   |||:::
908 CATTTAGCCATCTAACTTCACTTCATGAGTTGTACATGGGTCGTTGTAAI 957

108 ..............ProProAlaLeuArgGlu.................. 113
                  |||   :::|||::::::
958 CTGTCAGGGCCTATTCCTAAACCTCTATGGAATCTCACCAACATAGTGTT 1007

114 .LeuArgValSerGlyAsnAsnLeuThrSerLeuArgAlaLeuProSerG 130
     |||:::::::::::::|||:::|||   :::       :::       :::
1008 TTTGCACCTTGGTGATAACCATCTTGAAGGACCAATTTCCCATTTCACGA 1057

130 lyLeuGlnLysLeuTrpAlaTyrAsnAsnArgLeuThrSerLeuProGlu 146
     ::::::||||||            :::|||||||:::   :::   :::
1058 TATTTGAAAAGCTC............AAGAGGTTATCACTTGTAAATAAC 1095

147 ............................MetSerProGlyLeuGlnGl 153
                                      ::::::   |||:::
1096 AACTTTGATGGCGGACTTGAGTTCTTATCCTTTAACACCCAACTTGAACG 1145

153 uLeuAspValSerHisAsnGlnLeuThr...ArgLeuProGlnSerLeuT 169
     ||||||:::|||   |||   ||||||      :::|||   :::::::
1146 GCTAGATTTATCATCCAATTCCCTAACTGGTCCAATTCCATCCAACATAA 1195

169 hrGlyLeuSerSerAlaAlaArgValTyrLeuAspGlyAsnProLeuSer 185
    ::|||||   :::   :::   :::||||||:::::::|||   |||:::
1196 GCGGACTTCAAAACCTAGAATGTCTCTACTTGTCATCAAACCACTTGAAT 1245

186 ValArgThrArgAspArgLeuCysGlyHisHisTrpProPheArgHisGl 202

1246 GGG..................................................... 1248

202 nAspThrLeuArgTyrGlyGlyAlaPheArgProArgGluAlaArgAlaL 219

1248 ..................................................... 1248

219 euHisLeuAlaValAlaAspTrpLeuThrSerAlaArgGluGlyGluAla 235
           :::::::::::::|||:::
1249 ........AGTATACCTTCCTGGATA..................... 1266

236 AlaGlnAlaAspArgTrpGlnAlaPheGlyLeuGluAspAsnAlaAlaAl 252
                          |||:::|||
1267 ......................TTCTCCCTTCCT.............. 1278

252 aPheSerLeuVal...LeuAspArgLeuArgGluThrGluAsnPheLysL 268
     |||||||||   ||||||         :::|||
1279 ....TCACTGGTTGAGTTAGACTTGAGCAATAACACT............. 1311

268 ysAspAlaGlyPheLysAlaGlnIleSerSerTrpLeuThrGlnLeuAla 284
              |||   :::::::|||
1312 ..........TTCAGTGGAAAAATT...................... 1326

285 GluAspAlaAlaLeuArgAlaLysThrPheAlaMetAlaThrGluAlaTh 301
           :::::::::::::||||||::::::   :::|||
```

```
1327 ......CAAGAGTTCAAGTCCAAAACATTAAGTGCCGTTACTCTAAAACA 1370

 301 rSerThrCysGluAspArgValThrHisAlaLeuHisGlnMetAsnAsnV 318
1371 AAATAAGCTGAAAGGTCGTATTCCGAATTCACTCCTAAACCAGAAGAACC 1420

 318 alGlnLeuValHisAsnAlaGluLysGlyGluTyrAspAsnAsnLeuGln 334
1421 TACAATTACTTCTCCTTTCA..............CACAATAATATCAGT 1455

 335 GlyLeuValSerThrGlyArg...GluMetPheArgLeuAlaThrLeuGl 350
1456 GGACATATTTCTTCAGCTATCTGCAATCTGAAAACATTGATATTGTTAGA 1505

 350 uGlnIleAlaArgGluLysAlaGlyThrLeuAlaLeu...ValAspAspV 366
1506 CTTGGGAAGTAATAATTTGGAGGGAACAATCCCACAATGCGTGGTTGAGA 1555

 366 alGluValTyrLeuAlaPheGlnAsnLysLeuLysGluSerLeuGluLeu 382
1556 GGAACGAATACCTTTCGCATTTGGATTTGAGCAAAAACAGACTTAGTGGG 1605

 383 ThrSerValThrSerGluMetArgPhePheAspValSerGlyVal...Th 398
1606 ACAATCAATACAACT............TTTAGTGTTGGAAACATTTTAAG 1643

 398 rValSerAspLeuGlnAlaAlaAspValGlnValLysThrAlaGlu.... 413
1644 GGTCATTAGCTTGCACGGGAATAAGCTAACGGGGAAAGTCCCACGATCTA 1693

 413 ..................................................... 413

1694 TGATCAATTGCAAGTATTTGACACTACTTGATCTAGGTAACAATATGTTG 1743

 414 AsnSerGlyPheSerLysTrpIle.......................... 421
1744 AATGACACATTTCCAAACTGGTTGGGATACCTATTTCAATTGAAGATTTT 1793

 422 ....LeuGlnTrpGlyProLeuHisSer...................... 429
1794 AAGCTTGAGATCAAATAAGTTGCATGGTCCCATCAAATCTTCAGGGAATA 1843

 430 .......................ValLeuGlu.................. 432
1844 CAAACTTGTTTATGGGTCTTCAAATTCTTGATCTATCATCTAATGGATTT 1893

 433 ...ArgLysValProGluArg............PheAsnAlaLeuArgGl 444
1894 AGTGGGAATTTACCCGAAAGAATTTTGGGGAATTTGCAAACCATGAAGGA 1943

 444 uLysGln............IleSerAspTyr...GluAspThrTyrArgL 456
1944 AATTGATGAGAGTACAGGATTCCCAGAGTATATTTCTGATCCATATGATA 1993

 456 ysLeuTyrAspGluValLeuLysSerSerGlyLeuValAspAspThrAsp 472
1994 TTTATTACAATTATTTGACGACAATTTCTACAAAGGGACAAGATTATGAT 2043

 473 AlaGluArgThrIleGlyValSerAlaMetAspSerAlaLysLysGluPh 489
2044 TCTGTTCGAATTTTGGATTCTAACATGATTATCAATCTCTCAAAGAACAG 2093

 489 eLeuAspGly...Leu.ArgAlaLeuValAspGluValLeuGly...... 502
2094 ATTTGAAGGTCATATTCCAAGCATTATTGGAGATCTTGTTGGACTTCGTA 2143

 503 ......SerTyrLeuThrAlaArgTrpArgLeu 511
2144 CGTTGAACTTGTCTCACAATGTCTTGGAAGGTC 2176
```

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 1999, 17:14:12 ; Search time 1495.71 Seconds
(without alignments)
4537.970 Million cell updates/sec

Title: US-09-068-804-15
Perfect score: 3441
Sequence: 1 CGCAAAGCCGAGGAAACGAA.....AAACAGGCGAGTCTGGATGC 3441

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 986266752 residues

Database : EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	55.2	1.6	408	34	AA484400	AA484400 nf10e04.s
2	46	1.3	527	43	AI238960	AI238960 GH15115.5
3	42.8	1.2	686	23	D22789	D22789 RICC1332A R
4	42.6	1.2	620	51	AI728053	AI728053 BNLGHi969
c 5	41.4	1.2	554	51	AI728199	AI728199 BNLGHi101
c 6	41.2	1.2	628	51	AI731721	AI731721 BNLGHi105
c 7	40.6	1.2	590	44	AI292820	AI292820 GH15821.5
c 8	40.2	1.2	487	38	AA749517	AA749517 SWAMCAC06
9	40	1.2	323	30	AA208468	AA208468 mv85f02.r
c 10	39	1.1	650	33	AA390540	AA390540 LD08780.5
c 11	39	1.1	616	44	AI292805	AI292805 GH15806.5
c 12	39	1.1	468	44	AI292818	AI292818 GH15819.5
c 13	38.8	1.1	423	26	W54408	W54408 md02h05.r1
14	38.8	1.1	325	26	W57781	W57781 zd17b03.r1
c 15	38.8	1.1	513	29	AA174511	AA174511 ms69h04.r
c 16	38.8	1.1	445	34	AA475503	AA475503 vh15f12.r
c 17	38.8	1.1	555	35	AA545466	AA545466 vj99g12.r
c 18	38.8	1.1	369	43	AI197155	AI197155 ue51h03.r
c 19	38.6	1.1	531	38	AA441313	AA441313 LD16203.5
20	38	1.1	564	24	H76570	H76570 18275 Lambd
c 21	38	1.1	349	39	AA839883	AA839883 vw65d05.r
c 22	37.8	1.1	619	41	AI067558	AI067558 EST209240
23	37.8	1.1	420	43	AU033900	AU033900 AU033900
c 24	37.8	1.1	676	45	AI389106	AI389106 GH20192.5
25	37.6	1.1	237	23	R30281	R30281 12886 Lambd
26	37.4	1.1	558	34	AA483903	AA483903 ne92a10.s
c 27	37.4	1.1	580	37	AA699167	AA699167 HL07844.5
28	37.4	1.1	537	43	AI237995	AI237995 GH13939.5
29	37.4	1.1	557	46	AI404106	AI404106 GH23705.5
30	37.2	1.1	191	26	W88389	W88389 lmEST0337 L
c 31	37.2	1.1	292	30	AA260634	AA260634 va11f02.r
32	37.2	1.1	756	35	C25541	C25541 C25541 Dict
c 33	37.2	1.1	492	51	AI737278	AI737278 606038D02
34	37	1.1	314	28	AA083950	AA083950 zn16e01.r
35	37	1.1	363	42	AI083764	AI083764 qf23c06.x
c 36	36.8	1.1	337	41	AI021088	AI021088 ub01b01.r
37	36.8	1.1	609	51	AU070274	AU070274 AU070274
38	36.4	1.1	474	40	AA979220	AA979220 LD33551.5
39	36.4	1.1	736	44	AU039498	AU039498 AU039498
c 40	36.2	1.1	250	28	AA061257	AA061257 ml39c12.r
c 41	36.2	1.1	538	45	AA818804	AA818804 UI-R-A0-a
42	36	1.0	435	25	N79538	N79538 zb08h10.s1
43	36	1.0	382	28	AA119823	AA119823 mn14h06.r
c 44	36	1.0	601	47	AI484260	AI484260 EST248487
c 45	35.8	1.0	413	20	T32231	T32231 EST45490 Hu

ALIGNMENTS

RESULT 1
AA484400
LOCUS AA484400 408 bp mRNA EST 18-AUG-1997
DEFINITION nf10e04.s1 NCI_CGAP_Li1 Homo sapiens cDNA clone IMAGE:913374 similar to contains element MSR1 MSR1 repetitive element ;, mRNA sequence.
ACCESSION AA484400
NID g2213213

```
VERSION     AA484400.1  GI:2213213
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 408)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On May 9, 1995 this sequence version replaced gi:803122.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: David E. Kleiner, M.D., Ph.D., Rodrigo F.
            Chuaqui, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 638   Std Error: 0.00
            Seq primer: -41m13 fwd. ET from Amersham
            High quality sequence stop: 362.
FEATURES             Location/Qualifiers
     source          1. .408
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:913374"
                     /clone_lib="NCI_CGAP_Li1"
                     /tissue_type="liver"
                     /lab_host="DH10B"
                     /note="Vector: pAMP10; mRNA made from normal liver
                     hepatocytes, cDNA made by oligo-dT priming.
                     Non-directionally cloned. Size-selected on agarose gel,
                     average insert  size 600 bp. Reference:  Krizman et al.
                     (1996) Cancer Research 56:5380-5383."
BASE COUNT       64 a     17 c    208 g    119 t
ORIGIN
```

```
  Query Match             1.6%;  Score 55.2;  DB 34;  Length 408;
  Best Local Similarity   47.8%;  Pred. No. 0.00011;
  Matches  193;  Conservative    0;  Mismatches  208;  Indels    3;  Gaps    1;

Qy      46 GTCCTCGGCGCGCTGCTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTTACCGGTGGGGCG 105
           | | | |  | |  | |    || || |  |  ||||  |  ||| |    |  | | |
Db       5 GACATGGTGGTGGAGGTGGTGATGGTGATGGCGGTGGTGGTGGTTATGGTGGTGGAGATG 64

Qy     106 AGTCTGGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCG 165
           |||  | |||  |||| ||   | | ||| |||||||| |  | | | ||||  | |
Db      65 GTGGTGGTGATGGTGGCGGAGGTGGTGGTGGTGATGGTGGCAGTGGCAGTGGTGATAGTG 124

Qy     166 GCGACGGGAGTGTCGTTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAG 225
             |  ||       | | |||  |  || | |       | | |||| |   | | ||  |
Db     125 ATGGAGGTGACAATGGTCATTGTGGTGGTGGTTGTTGTGGTAATGGTGGTGGAGGTGGTG 184

Qy     226 CCGTTAATGGAGCTGATTGGCAAGGCGATTACCAAAGCGCTGGAAGGATTAGGCGTCGAT 285
             | | ||||   || | |    || |||    |  |   |||  | | | |   | |
Db     185 ATGGTGATGGTAGTGGTGGCAGTGGTGATGGTGATGGTAGTGGTGGCAGTGGTGATAGTG 244

Qy     286 AAGAAAACGGCAGAGATGGCCGGCAGCATTGTTGGTGCGATTGTCGCCGCTATTGCCATG 345
           | | |   ||    | ||  |    | | |  | ||| ||   |  | || ||
Db     245 ATGGAGGTGGTGATGGTGATGGCGGCAGTGGCAGTGGTGATAGTGATGGAGGTGGCAATA 304

Qy     346 GTG---GCGGTCATTGTGGTGGTCGCAGTTGTCGGGAAAGGCGCGGCGGCGAAACTGGGT 402
           ||    | | | || |||||||| | | || | | | |  | |  || ||   | |
Db     305 GTCATTGTGATGATGGTGGTGGTGGTAATGGTGGTGGAGGTGGTGATGGTGATGGTAGTG 364

Qy     403 AACGCGCTGAGCAAAATGATGGGCGAAACGATTAAGAAGTTGGT 446
             ||  ||   | | ||||||  |  | |||    | | ||||
Db     365 GTGGCAGTGGTGATAGTGATGGAGGTGATGATGGTCATGGTGGT 408
```

```
RESULT   2
AI238960
LOCUS       AI238960      527 bp    mRNA            EST       18-NOV-1998
DEFINITION  GH15115.5prime GH Drosophila melanogaster head pOT2 Drosophila
            melanogaster cDNA clone GH15115 5prime, mRNA sequence.
ACCESSION   AI238960
NID         g3833818
VERSION     AI238960.1  GI:3833818
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 527)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     On Aug 21, 1998 this sequence version replaced.

            Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Sequence is the complete cDNA Insert.
            Plate: 151  row: B  column: 3
            High quality sequence stop: 343.
FEATURES             Location/Qualifiers
     source          1. .527
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="GH15115"
                     /clone_lib="GH Drosophila melanogaster head pOT2"
                     /sex="male and female"
                     /dev_stage="adult"
                     /lab_host="DH5 - alpha"
                     /note="Organ: head; Vector: pOT2; Site_1: EcoRI; Site_2:
                     XhoI; Sized fractionated cDNAs were directly ligated into
                     pOT2.  Plasmid cDNA library."
BASE COUNT      131 a    108 c    166 g    122 t
ORIGIN
```

```
  Query Match             1.3%;  Score 46;  DB 43;  Length 527;
  Best Local Similarity   49.6%;  Pred. No. 0.044;
  Matches  118;  Conservative    0;  Mismatches  120;  Indels    0;  Gaps    0;

Qy     155 TTGTGAAGGCGGCGACGGGAGTGTCGTTTATTCAGCAGGCGCIAAACCCGATTATGGAGC 214
           ||  |  ||||| | |||  ||| ||    |    ||| | |   |   | ||||||
Db      84 TTCCGCCGGCGGTGGCGGTGGTGGCGGAGGTGGCCAAGGAGGTTGGCAGAAGAATGGAGG 143

Qy     215 ATGTGCTGAAGCCGTTAATGGAGCTGATTGGCAAGGCGATTACCAAAGCGCTGGAAGGAT 274
           | | |  | ||  | |  |  || |  ||||||   |    |    |||  ||| |
Db     144 AGGAGGAGGAGGAGGTGGTCAAGGAGGCTGGCAAAAGGGCGGCGGCGGCGGAGGAGGTGG 203

Qy     275 TAGGCGTCGATAAGAAAACGGCAGAGATGGCCGGCAGCATTGTTGGTGCGATTGTCGCCG 334
           |  | | ||    |    |||  |||  ||||  |||     || |   ||   || | |
Db     204 CAAACATGGAGGTGGTGGCGGAGGAGGTGGCAAGCATGGAGGTGGAAACGGCGGTGGTGG 263

Qy     335 CTATTGCCATGGTGGCGGTCATTGTGGTGGTCGCAGTTGTCGGGAAAGGCGCGGCGGC 392
           |   |  ||| ||||||    | |||||  |  ||  |   | |  |  ||||||||
Db     264 AAAGCACGGTGGCGGCGGTGGAGGCGGTGGAGGTGGTGGAAAGCATGGAGGCGGCGGC 321
```

```
RESULT   3
```

```
D22789
LOCUS       D22789         686 bp    mRNA            EST       20-JUL-1998
DEFINITION  RICC1332A Rice callus Oryza sativa cDNA clone R, mRNA sequence.
ACCESSION   D22789
NID         g871921
VERSION     D22789.1  GI:871921
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1  (bases 1 to 686)
  AUTHORS   Sasaki,T. and Minobe,Y.
  TITLE     Rice cDNA from callus
  JOURNAL   Unpublished (1994)
COMMENT     On Dec 2, 1993 this sequence version replaced.

            Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program
            2-1-2 Kannondai,Tsukuba
            Ibaraki,
            Japan 305
            Tel: 0298-38-7441
            Fax: 0298-38-7468
            Email: tsasaki@abr.affrc.go.jp
            PROJECT ='RGP'.
FEATURES             Location/Qualifiers
     source          1. .686
                     /organism="Oryza sativa"
                     /strain="cultivar Nipponbare, sub_species Japonica"
                     /db_xref="taxon:4530"
                     /map="2"
                     /clone="R"
                     /clone_lib="Rice callus"
                     /note="Vector: pBluescript II SK+; Site_1: SalI; Site_2:
                     NotI; cDNA prepared from rice callus mRNAs by using
                     oligo(dT) as a primer and ligating to the SalI-NotI site
                     of pBluescript II SK+ phagemid.  "
BASE COUNT      125 a    132 c    246 g    183 t
ORIGIN
```

```
  Query Match             1.2%;  Score 42.8;  DB 23;  Length 686;
  Best Local Similarity   46.6%;  Pred. No. 0.37;
  Matches  137;  Conservative    0;  Mismatches  157;  Indels    0;  Gaps    0;

Qy     100 GGGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTG 159
           | || |    ||| | |||    || ||   | | ||| | |||||     |  || ||
Db     108 GTGGAGGTGGTGGTGGTGGTGAAGGAGGCGGTGGTGGTTACGGTGGATCGGGCTATGGTT 167

Qy     160 AAGGCGGCGACGGGAGTGTCGTTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTG 219
             ||     |     |||  |    |   | ||| ||      |  | |||   | |
Db     168 CTGGTTCTGGGTATGGTGAGGGAGGTGGCTCCGGCTCTGGCTATGGGTCTGGGCAAGGCT 227

Qy     220 CTGAAGCCGTTAATGGAGCTGATTGGCAAGGCGATTACCAAAGCGCTGGAAGGATTAGGC 279
           |||     | |   || | || | |    ||||  ||    || | |||  |   | |
Db     228 CTGGCTATGGTGCAGGTGTTGGTGGTGCTGGCGGGTATGGTAGTGGTGGCGGAGGTGGTG 287

Qy     280 GTCGATAAGAAAACGGCAGAGATGGCCGGCAGCATTGTTGGTGCGATTGTCGCCGCTATT 339
           || |  |||     ||    |  |||     |||  | |   || ||  | |||||
Db     288 GTGGCCAAGGTGGTGGTGCTGGCGGCTACGGGCAAGGGTCTGGCTATGGTTCCGGCTATG 347

Qy     340 GCCATGGTGGCGGTCATTGTGGTGGTCGCAGTTGTCGGGAAAGGCGCGGCGGCG 393
           |    ||||  |||  |  |  ||||  |  |||  | |     |  ||||  ||||
Db     348 GGTCGGGTGCTGGTGGTGCTCATGGTGGTGGTTATGGAAGTGGTGGCGGTGGCG 401
```

```
RESULT   4
AI728053
LOCUS       AI728053       620 bp    mRNA            EST       11-JUN-1999
DEFINITION  BNLGHi9696 Six-day Cotton fiber Gossypium hirsutum cDNA 5' similar
            to (AF051898) coronin binding protein [Dictyostelium discoideum],
            mRNA sequence.
ACCESSION   AI728053
NID         g5046839
VERSION     AI728053.1  GI:5046839
KEYWORDS    EST.
SOURCE      upland cotton.
  ORGANISM  Gossypium hirsutum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
REFERENCE   1  (bases 1 to 620)
  AUTHORS   Blewitt,M., Matz,E.C., Davy,D.F. and Burr,B.
  TITLE     ESTs from developing cotton fiber
  JOURNAL   Unpublished (1999)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3187769.

            Contact: Ben Burr
            Biology Department
            Brookhaven National Laboratory
            Upton, NY 11973, USA
            Tel: 516-344-3396
            Fax: 516-344-3407
            Email: burr@bnlux1.bnl.gov
            Seq primer: T3 Primer.
FEATURES             Location/Qualifiers
     source          1. .620
                     /organism="Gossypium hirsutum"
                     /cultivar="Acala Maxxa"
                     /db_xref="taxon:3635"
                     /map="17q21; 8q13.2"
                     /clone_lib="Six-day Cotton fiber"
                     /tissue_type="immature fiber"
                     /dev_stage="Six days post anthesis"
                     /lab_host="XL1-Blue"
                     /note="Vector: pBluescript II KS+"
BASE COUNT      259 a    157 c    122 g     82 t
ORIGIN
```

```
  Query Match             1.2%;  Score 42.6;  DB 51;  Length 620;
  Best Local Similarity   47.9%;  Pred. No. 0.41;
  Matches  123;  Conservative    0;  Mismatches  134;  Indels    0;  Gaps    0;

Qy    3180 GTACGATAAGTAGCAGCCTGGAAACCGCCAAAAGCTTCCTGCAAGGATAACAGAAGAGGA 3239
           | || || || |   |  |   |||   || ||   | || |||| |  || ||  |  |
Db       1 GAACAATCAGAATGCGGATCAGAACGAGCACAATGCTGCTCCAAGTAACACTGAGAACAA 60

Qy    3240 TATTAATAATGGTTACAAGTGTAAGGACTCAGCCCCCCGTCATAATGCCAGGTATGCAGA 3299
           |   |||   ||| | ||    ||  |  |  | ||  || | |  | ||   |||   |
Db      61 TGAGAATGCGGGTCAGAACGAGAACAATGCCACTCCAAGTAACACCGACAACAATGAGAA 120

Qy    3300 CCGAGATCAAAACGCAGGCCACGAATCTTGCGGCGAATCTTTCCGCAGTCAGAGAAAGTG 3359
             || ||| ||||   | | |   ||    || |  ||  |  |||  | || ||
Db     121 TGCAGGTCAGAACGAGAACAATGTGACTCAGAGCAACACTGACAACAGCAACAGCAACAA 180

Qy    3360 CCACAGCGACGCTGTCAGGGGAAATTAAAGGCCCGCAACTGGAAGATTTTCCCGCGCTGA 3419
           | ||| |||   ||  ||   |||  | |     ||  ||   ||  |  |  |  |||
Db     181 CAACAACGAAAATGCAAGTCAAAACGAGAACAATGCTGCTCAGAGCAGTAACAACAATGA 240

Qy    3420 TCAAACAGGCGAGTCTG 3436
           |  || |  ||  ||
Db     241 AAATGCAAGTCAGGATG 257
```

```
RESULT   5
AI728199/c
LOCUS       AI728199       554 bp    mRNA            EST       11-JUN-1999
DEFINITION  BNLGHi10138 Six-day Cotton fiber Gossypium hirsutum cDNA 5' similar
            to (AL022537) putative protein [Arabidopsis thaliana], mRNA
            sequence.
ACCESSION   AI728199
NID         g5047051
```

```
VERSION     AI728199.1  GI:5047051
KEYWORDS    EST.
SOURCE      upland cotton.
  ORGANISM  Gossypium hirsutum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
REFERENCE   1  (bases 1 to 554)
  AUTHORS   Blewitt,M., Matz,E.C., Davy,D.F. and Burr,B.
  TITLE     ESTs from developing cotton fiber
  JOURNAL   Unpublished (1999)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3187982.

            Contact: Ben Burr
            Biology Department
            Brookhaven National Laboratory
            Upton, NY 11973, USA
            Tel: 516-344-3396
            Fax: 516-344-3407
            Email: burr@bnlux1.bnl.gov
            Seq primer: T3 Primer.
FEATURES             Location/Qualifiers
     source          1. .554
                     /organism="Gossypium hirsutum"
                     /cultivar="Acala Maxxa"
                     /db_xref="taxon:3635"
                     /clone_lib="Six-day Cotton fiber"
                     /tissue_type="immature fiber"
                     /dev_stage="Six days post anthesis"
                     /lab_host="XL1-Blue"
                     /note="Vector: pBluescript II KS+"
BASE COUNT      109 a    226 c     64 g    155 t
ORIGIN
```

```
  Query Match             1.2%;  Score 41.4;  DB 51;  Length 554;
  Best Local Similarity   49.5%;  Pred. No. 0.83;
  Matches  136;  Conservative    0;  Mismatches  136;  Indels    3;  Gaps    1;

Qy    101 GGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGA 160
          ||  ||  | |  || ||  |||  |||| | | || | ||| ||  ||  |  |  | ||
Db    351 GGAAGAAGCCGAGGCGGGAAGCGTCGGGATTGGGGGCATTGGCGGAGGAAGTGACCGAGA 292

Qy    161 AGGCGGCGACGGGAGTGTCGTTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGC 220
           || || |||||   || || |    | | ||  | | || |||| | |   | |    |
Db    291 CGGAGGAGACGGAGATGGCGATGTGACCGGAGTGGGTGAAGCCGAATTTCTCGAAGAGGA 232

Qy    221 TGAAGCCGTTAAT---GGAGCTGATTGGCAAGGCGATTACCAAAGCGCTGGAAGGATTAG 277
          | | |  |  |     |||| |||| |  | ||   | | |   ||| |||| |    |
Db    231 TCATGGGGCGAGAGTCGGAGGTGATGGTTAGGGTTTTGATCTCTGCGGTGGAGGAGGAAA 172

Qy    278 GCGTCGATAAGAAAACGGCAGAGATGGCCGGCAGCATTGTTGGTGCGATTGTCGCCGCTA 337
             || | ||  | |     || |  | | |     |  |    || ||| | |  |
Db    171 TGAGGGAAAGGAGGAAGAGGAGGAGGATCTGTAAGGGTAGTTTCGCCATTTTGGGAGAAT 112

Qy    338 TTGCCATGGTGGCGGTCATTGTGGTGGTCGCAGTT 372
          ||  |   |||||      |  || |  || | ||||
Db    111 TTACGGCGGTGGTTACGACGGTCGGCGTGGTAGTT 77
```

```
RESULT   6
AI731721/c
LOCUS       AI731721      628 bp    mRNA            EST       11-JUN-1999
DEFINITION  BNLGHi10580 Six-day Cotton fiber Gossypium hirsutum cDNA 5' similar
            to (AL031032) extensin-like protein [Arabidopsis thaliana], mRNA
            sequence.
ACCESSION   AI731721
NID         g5050573
VERSION     AI731721.1  GI:5050573
KEYWORDS    EST.
SOURCE      upland cotton.
  ORGANISM  Gossypium hirsutum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
REFERENCE   1  (bases 1 to 628)
  AUTHORS   Blewitt,M., Matz,E.C., Davy,D.F. and Burr,B.
  TITLE     ESTs from developing cotton fiber
  JOURNAL   Unpublished (1999)
COMMENT     On Mar 10, 1998 this sequence version replaced gi:2948508.

            Contact: Ben Burr
            Biology Department
            Brookhaven National Laboratory
            Upton, NY 11973, USA
            Tel: 516-344-3396
            Fax: 516-344-3407
            Email: burr@bnlux1.bnl.gov
            Seq primer: T3 Primer.
FEATURES             Location/Qualifiers
     source          1. .628
                     /organism="Gossypium hirsutum"
                     /cultivar="Acala Maxxa"
                     /db_xref="taxon:3635"
                     /clone_lib="Six-day Cotton fiber"
                     /tissue_type="immature fiber"
                     /dev_stage="Six days post anthesis"
                     /lab_host="XL1-Blue"
                     /note="Vector: pBluescript II KS+"
BASE COUNT      144 a    306 c     44 g    134 t
ORIGIN
```

```
  Query Match             1.2%;  Score 41.2;  DB 51;  Length 628;
  Best Local Similarity   44.2%;  Pred. No. 1;
  Matches  169;  Conservative    0;  Mismatches  213;  Indels    0;  Gaps    0;

Qy     97 GGTGGGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATT 156
          || | |  |   |||  | |||  |      | | | | |||   ||||   |      |
Db    474 GGGGTGAAGGTGATGGAGGTGGTTTGTAATGAATTGGGGGTGGGGGTGATGGTGATGGTG 415

Qy    157 GTGAAGGCGGCGACGGGAGTGTCGTTTATTCAGCAGGCGCTAAACCCGATTATGGAGCAT 216
          |||  || || |  |   | ||    | | ||   || |||      | |  ||| | |
Db    414 GTGGGGGTGGGGGTGATGGCGTGTACTCTACACACGGTGCTGGTGGTGGTGGTGGTGGAG 355

Qy    217 GTGCTGAAGCCGTTAATGGAGCTGATTGGCAAGGCGATTACCAAAGCGCTGGAAGGATTA 276
          |||    |      |   |||| |||  |    || | | |   ||  |  |||  |
Db    354 GTGGCTCAATACATGGAGGAGGTGACGGTGGTGGTGGTGAGTGAACTGGAGGAGGTGGGG 295

Qy    277 GGCGTCGATAAGAAAACGGCAGAGATGGCCGGCAGCATTGTTGGTGCGATTGTCGCCGCT 336
             |  ||     ||| | | |||     || || | |     | | |   |  |
Db    294 AATAAACAGGAGGTGGTGGCGGTGGTGGAGATAAGTATGGATAAATTGGTGGAGGAGGTG 235

Qy    337 ATTGCCATGGTGGCGGTCATTGTGGTGGTCGCAGTTGTCGGGAAAGGCGCGGCGGCGAAA 396
          | ||   ||||| |||       ||     | || |  | ||| || | ||||| |
Db    234 AATGTGGTGGTGGTGGTGGAGAATGTACCGGAGGTGGCGGTGAATGGTGTGGCGGTGGCG 175

Qy    397 CTGGGTAACGCGCTGAGCAAAATGATGGGCGAAACGATTAAGAAGTTGGTGCCTAACGTG 456
          | || |       |   || || |||  |    ||  |  | |  ||||    | | |
Db    174 GGGAGTTATAGTAGTAAGGAACTGGTGGAGGTGGGGAAAACAATGGAGGTGGTGGAGGGG 115

Qy    457 CTGAAACAGTTGGCGCAAAACG 478
          ||||   | ||| | |  | |
Db    114 GTGAATTTGGTGGTGGAGGAGG 93
```

```
RESULT   7
AI292820/c
LOCUS       AI292820      590 bp    mRNA            EST       01-DEC-1998
DEFINITION  GH15821.5prime GH Drosophila melanogaster head pOT2 Drosophila
            melanogaster cDNA clone GH15821 5prime, mRNA sequence.
ACCESSION   AI292820
NID         g3942227
VERSION     AI292820.1  GI:3942227
KEYWORDS    EST.
SOURCE      fruit fly.
```

```
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 590)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     On Apr 14, 1993 this sequence version replaced gi:639346.

            Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            hit genomic sequence AL023893
            Plate: 158  row: B  column: 9
            High quality sequence stop: 433.
FEATURES             Location/Qualifiers
     source          1. .590
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="GH15821"
                     /clone_lib="GH Drosophila melanogaster head pOT2"
                     /sex="male and female"
                     /dev_stage="adult"
                     /lab_host="DH5 - alpha"
                     /note="Organ: head; Vector: pOT2; Site_1: EcoRI; Site_2:
                     XhoI; Sized fractionated cDNAs were directly ligated into
                     pOT2.  Plasmid cDNA library."
BASE COUNT       172 a    191 c    132 g     95 t
ORIGIN
```

```
  Query Match             1.2%;  Score 40.6;  DB 44;  Length 590;
  Best Local Similarity   49.3%;  Pred. No. 1.4;
  Matches  106;  Conservative    0;  Mismatches  109;  Indels    0;  Gaps    0;

Qy      42 GAAAGTCCTCGGCGCGCTGCTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTGG 101
           | | ||  |  |   || ||   ||  |||    || || |  | |||| ||
Db     281 GCATGTGGTGCGGATGCGGCCCGCCCATGTTGTGGTGGTTGTTGTTGTTGTTGTTATGAT 222

Qy     102 GGCGAGTCTGGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAA 161
           || |  | | | | ||  ||  || |   |||  |  |||||||       |  | ||| |
Db     221 GGTGGTTGTTGGGGTGATTGTTGTTGTGGTTGTTGGGATGGTGGTGCTGCATGTGGTGCA 162

Qy     162 GGCGGCGACGGGAGTGTCGTTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCT 221
           |  |  || ||   ||  | |  |   |  |  |||    | ||| |||  ||    ||
Db     161 GCTGCGGATGGTGCTGCGGATGCTGGTGTGGATGCTGCGGCTGATGATGATGCGGCGGCA 102

Qy     222 GAAGCCGTTAATGGAGCTGATTGGCAAGGCGATTA 256
           | || | |  ||| | |  |||  | || |||||
Db     101 TATGCGGATTGTGGTGGTTGTTGTGATGGTGATTA 67
```

```
RESULT   8
AA749517/c
LOCUS       AA749517      487 bp    mRNA            EST       20-JAN-1998
DEFINITION  SWAMCAC06A12SK Brugia malayi adult male cDNA (SAW94NL-BmAM) Brugia
            malayi cDNA clone SWAMCAC06A12 5', mRNA sequence.
ACCESSION   AA749517
NID         g2796223
VERSION     AA749517.1  GI:2796223
KEYWORDS    EST.
SOURCE      Brugia malayi.
  ORGANISM  Brugia malayi
            Eukaryota; Metazoa; Nematoda; Secernentea; Spiruria; Spirurida;
            Filarioidea; Onchocercidae; Brugia.
REFERENCE   1  (bases 1 to 487)
  AUTHORS   Williams,S.A.
  TITLE     Genes expressed in adult males of Brugia malayi
  JOURNAL   Unpublished (1995)
COMMENT     On Jan 9, 1998 this sequence version replaced gi:693559.

            Contact: Steven A. Williams
            Molecular Parasitology
            Smith College Department of Biological Sciences
            Department of Biological Sciences, Clark Science Center, Smith
            College, Northampton, MA, 01063, USA
            Tel: 4135853826
            Fax: 4135853786
            Email: genome@smith.edu
            Seq primer: pBluescript SK.
FEATURES             Location/Qualifiers
     source          1. .487
                     /organism="Brugia malayi"
                     /strain="TRS Labs"
                     /db_xref="taxon:6279"
                     /clone="SWAMCAC06A12"
                     /clone_lib="Brugia malayi adult male cDNA (SAW94NL-BmAM)"
                     /lab_host="XL1-Blue MRF'"
                     /note="Vector: lambda UniZap XR; Site_1: EcoR I; Site_2:
                     Xho I; Lymphatic filarial nematode parasite of humans.
                     mRNA was prepared from adult males of Brugia malayi
                     isolated from jirds and converted to double stranded cDNA
                     using reverse transcriptase and oligo(dT) followed by
                     RNase H and DNApol I. The library had 4.6 x 10E6
                     independent recombinants and average insert size was 800
                     base pairs.  The library was constructed by Noelle Ling.
                     The library is available from Dr. S.A. Williams, email
                     genome@smith.edu."
BASE COUNT       138 a    175 c     71 g    102 t      1 others
ORIGIN
```

```
  Query Match             1.2%;  Score 40.2;  DB 38;  Length 487;
  Best Local Similarity   48.9%;  Pred. No. 1.7;
  Matches  108;  Conservative    0;  Mismatches  113;  Indels    0;  Gaps    0;

Qy      37 ATCGGGAAAGTCCTCGGCGCGCTGCTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACC 96
           || | ||  ||  | |  |  | |  |    |  | ||  | ||| | |  | | ||
Db     256 ATGGAGATGGTGGTGGTGGTTCAGGCACTGGTGATGAGGATAGTGACGGTGGCTGTTCGG 197

Qy      97 GGTGGGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATT 156
           |    |||||    | |  |||| |  | ||   | | ||||||||||||  |  | | |
Db     196 GTATTGGCGATGACGACTGTGGCAGTAGCGGTGGTGGTGGTAATGGTGGTGGTGGTAGTG 137

Qy     157 GTGAAGGCGGCGACGGGAGTGTCGTTTATTCAGCAGGCGCTAAACCCGATTATGGAGCAT 216
           ||   || || | |||    |  | | ||   |  ||  | |       |   || ||| |
Db     136 GTACCGGTGGTGGCGGTTCAGATGCTAATGATGGAGTTGGTGGCAATGGCGATAGAGAAA 77

Qy     217 GTGCTGAAGCCGTTAATGGAGCTGATTGGCAAGGCGATTAC 257
           | | ||  |  | |    | | |||||    ||| | | ||
Db      76 GGGATGCGGGAGGTGGGAGTGATGATTCAACAGGTGCTGAC 36
```

```
RESULT   9
AA208468
LOCUS       AA208468      323 bp    mRNA            EST       12-MAR-1997
DEFINITION  mv85f02.r1 GuayWoodford Beier mouse kidney day 7 Mus musculus cDNA
            clone IMAGE:661851 5', mRNA sequence.
ACCESSION   AA208468
NID         g1804982
VERSION     AA208468.1  GI:1804982
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 323)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
```

```
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1393272.

            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:407699
            Putative full length read
            vector to vector length is 515
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 306.
FEATURES             Location/Qualifiers
     source          1. .323
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:661851"
                     /clone_lib="GuayWoodford Beier mouse kidney day 7"
                     /tissue_type="kidney"
                     /dev_stage="juvenile (7 days old)"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Organ: kidney; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. Average insert size: 1.0 kb; Uni-ZAP XR Vector;
                     ~5' adaptor sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor
                     sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3' Library provided
                     by Lisa Guay-Woodford."
BASE COUNT      128 a     58 c     52 g     85 t
ORIGIN


  Query Match             1.2%;  Score 40;  DB 30;  Length 323;
  Best Local Similarity   57.0%;  Pred. No. 1.6;
  Matches   73;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy    2673 TCTATACCGATTTTTATCAGGCCTTCAGTGATATTCTTTCCAAAATGGGAGGCTGGTTAT 2732
           |||  |   |  ||||  |      ||| | | | |      ||||| | |  ||  |||
Db     159 TCTGAAAATACATTTAAAAAAAACTCATTTAGAGTTAGGGAAAAATAGAACCCTTATTAG 218

Qy    2733 TACCAGGTAAGGACGGTAATACCGTTAAGCTAGATGTTACCTCACTCAAAAATGATTTAA 2792
           || |   | |  || |  | |||  ||||  |||   ||  ||||  ||||| | |||||
Db     219 TAACCAATCATTACAGACAAACCAATAAGGAAGAAACTAGGTCACCTAAAAAAGTTTTAA 278

Qy    2793 ACAGTTTA 2800
           ||||||||
Db     279 ACAGTTTA 286


RESULT  10
AA390540/c
LOCUS       AA390540      650 bp    mRNA            EST       28-NOV-1998
DEFINITION  LD08780.5prime LD Drosophila melanogaster embryo BlueScript
            Drosophila melanogaster cDNA clone LD08780 5prime, mRNA sequence.
ACCESSION   AA390540
NID         g2043534
VERSION     AA390540.1  GI:2043534
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 650)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1395282.

            Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 87  row: G  column: 8
            High quality sequence stop: 482.
FEATURES             Location/Qualifiers
     source          1. .650
                     /organism="Drosophila melanogaster"
                     /db_xref="BDGP_EST:BDcln008107"
                     /db_xref="taxon:7227"
                     /clone="LD08780"
                     /clone_lib="LD Drosophila melanogaster embryo BlueScript"
                     /sex="male and female"
                     /dev_stage="0 to 24 hours mixed stage embryonic"
                     /lab_host="SOLR"
                     /note="Organ: embryo; Vector: BlueScript SK; Site_1:
                     EcoRI; Site_2: XhoI; Constructed using Stratagene ZAP-cDNA
                     Synthesis kit. Oligo dT-primed and directionally cloned at
                     EcoRI and XhoI in BlueScript SK(+/-)"
BASE COUNT      182 a    216 c    158 g     94 t
ORIGIN


  Query Match             1.1%;  Score 39;  DB 33;  Length 650;
  Best Local Similarity   48.8%;  Pred. No. 4.1;
  Matches  105;  Conservative    0;  Mismatches  110;  Indels    0;  Gaps    0;

Qy      42 GAAAGTCCTCGGCGCGCTGCTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTGG 101
           | | ||  |  |   || |    ||  |||    || || |  | |||| ||
Db     579 GCATGTGGTGCGGATGCGGGCCGCCCATGTTGTGGTGGTTGTTGTTGTTGTTGTTATGAT 520

Qy     102 GGCGAGTCTGGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAA 161
           || |  | | | | ||  ||  || |   |||  |  |||||||      |  | ||| |
Db     519 GGTGGTTGTTGGGGTGATTGTTGTTGTGGTTGTTGGGATGGTGGTGCTGCATGTGGTGCA 460

Qy     162 GGCGGCGACGGGAGTGTCGTTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCT 221
           |  |  || ||   ||  | |  |   |  | |||    | ||| ||| ||   ||
Db     459 GCTGCGGATGGTGCTGCGGATGCTGGTGTGGATGCTGCGGCTGATGATGATGCGGCGGCA 400

Qy     222 GAAGCCGTTAATGGAGCTGATTGGCAAGGCGATTA 256
           | || | |  ||| | |  |||  | || |||||
Db     399 TATGCGGATTGTGGTGGTTGTTGTGATGGTGATTA 365


RESULT  11
AI292805/c
LOCUS       AI292805      616 bp    mRNA            EST       01-DEC-1998
DEFINITION  GH15806.5prime GH Drosophila melanogaster head pOT2 Drosophila
            melanogaster cDNA clone GH15806 5prime, mRNA sequence.
ACCESSION   AI292805
NID         g3942212
VERSION     AI292805.1  GI:3942212
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 616)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     On May 1, 1998 this sequence version replaced gi:2761092.

            Contact: Harvey, D.
```

```
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            hit genomic sequence AL023893
            Plate: 158  row: A  column: 6
            High quality sequence stop: 544.
FEATURES              Location/Qualifiers
     source           1. .616
                      /organism="Drosophila melanogaster"
                      /db_xref="taxon:7227"
                      /clone="GH15806"
                      /clone_lib="GH Drosophila melanogaster head pOT2"
                      /sex="male and female"
                      /dev_stage="adult"
                      /lab_host="DH5 - alpha"
                      /note="Organ: head; Vector: pOT2; Site_1: EcoRI; Site_2:
                      XhoI; Sized fractionated cDNAs were directly ligated into
                      pOT2.  Plasmid cDNA library."
BASE COUNT      178 a    202 c    138 g     98 t
ORIGIN


  Query Match             1.1%;  Score 39;  DB 44;  Length 616;
  Best Local Similarity   48.8%;  Pred. No. 4;
  Matches  105;  Conservative    0;  Mismatches  110;  Indels    0;  Gaps    0;

Qy      42 GAAAGTCCTCGGCGCGCTGCTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTGG 101
           | | ||  |  |   || |    ||  |||    || || |  | |||| ||
Db     281 GCATGTGGTGCGGATGCGGGCCGCCCATGTTGTGGTGGTTGTTGTTGTTGTTGTTATGAT 222

Qy     102 GGCGAGTCTGGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAA 161
           || |  | | | | ||  ||  || |   |||  |  |||||||      |  | ||| |
Db     221 GGTGGTTGTTGGGGTGATTGTTGTTGTGGTTGTTGGGATGGTGGTGCTGCATGTGGTGCA 162

Qy     162 GGCGGCGACGGGAGTGTCGTTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCT 221
           |  |  || ||   ||  | |  |   |  |  |||    | ||| |||  ||    ||
Db     161 GCTGCGGATGGTGCTGCGGATGCTGGTGTGGATGCTGCGGCTGATGATGATGCGGCGGCA 102

Qy     222 GAAGCCGTTAATGGAGCTGATTGGCAAGGCGATTA 256
           | || | |  ||| | |  |||  | || |||||
Db     101 TATGCGGATTGTGGTGGTTGTTGTGATGGTGATTA 67


RESULT  12
AI292818/c
LOCUS       AI292818      468 bp    mRNA            EST       01-DEC-1998
DEFINITION  GH15819.5prime GH Drosophila melanogaster head pOT2 Drosophila
            melanogaster cDNA clone GH15819 5prime, mRNA sequence.
ACCESSION   AI292818
NID         g3942225
VERSION     AI292818.1  GI:3942225
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 468)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     On Apr 14, 1993 this sequence version replaced gi:639324.

            Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            hit genomic sequence AL023893
            Plate: 158  row: B  column: 7
            High quality sequence stop: 433.
FEATURES              Location/Qualifiers
     source           1. .468
                      /organism="Drosophila melanogaster"
                      /db_xref="taxon:7227"
                      /clone="GH15819"
                      /clone_lib="GH Drosophila melanogaster head pOT2"
                      /sex="male and female"
                      /dev_stage="adult"
                      /lab_host="DH5 - alpha"
                      /note="Organ: head; Vector: pOT2; Site_1: EcoRI; Site_2:
                      XhoI; Sized fractionated cDNAs were directly ligated into
                      pOT2.  Plasmid cDNA library."
BASE COUNT      139 a    161 c    109 g     59 t
ORIGIN


  Query Match             1.1%;  Score 39;  DB 44;  Length 468;
  Best Local Similarity   48.8%;  Pred. No. 3.6;
  Matches  105;  Conservative    0;  Mismatches  110;  Indels    0;  Gaps    0;

Qy      42 GAAAGTCCTCGGCGCGCTGCTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTGG 101
           | | ||  |  |   || |    ||  |||    || || |  | |||| ||
Db     281 GCATGTGGTGCGGATGCGGGCCGCCCATGTTGTGGTGGTTGTTGTTGTTGTTGTTATGAT 222

Qy     102 GGCGAGTCTGGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAA 161
           || |  | | | | ||  ||  || |   |||  |  |||||||      |  | ||| |
Db     221 GGTGGTTGTTGGGGTGATTGTTGTTGTGGTTGTTGGGATGGTGGTGCTGCATGTGGTGCA 162

Qy     162 GGCGGCGACGGGAGTGTCGTTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCT 221
           |  |  || ||   ||  | |  |   |  |  |||    | ||| |||  ||    ||
Db     161 GCTGCGGATGGTGCTGCGGATGCTGGTGTGGATGCTGCGGCTGATGATGATGCGGCGGCA 102

Qy     222 GAAGCCGTTAATGGAGCTGATTGGCAAGGCGATTA 256
           | || | |  ||| | |  |||  | || |||||
Db     101 TATGCGGATTGTGGTGGTTGTTGTGATGGTGATTA 67


RESULT  13
W54408/c
LOCUS       W54408        423 bp    mRNA            EST       03-JUN-1996
DEFINITION  md02h05.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
            clone IMAGE:367257 5', mRNA sequence.
ACCESSION   W54408
NID         g1355450
VERSION     W54408.1  GI:1355450
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 423)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Apr 14, 1993 this sequence version replaced gi:692592.

            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:228689
            Seq primer: ETPrimer
```

```
            High quality sequence stop: 345.
FEATURES             Location/Qualifiers
     source          1. .423
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /map="7"
                     /clone="IMAGE:367257"
                     /clone_lib="Soares mouse embryo NbME13.5 14.5"
                     /sex="unknown"
                     /tissue_type="embryo"
                     /dev_stage="13.5-14.5dpc total fetus"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGGAAATTTTTTTTTTTTTTTTTTTTTT
                     T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
                     14.5dpc embryos [total RNA provided by  Minoru Ko, Wayne
                     State Univ., from 2 ]; double-stranded cDNA was ligated to
                     Eco RI adaptors (Pharmacia), digested with Not I and
                     cloned into the Not I and Eco RI sites of the modified
                     pT7T3 vector. Library went through one round of
                     normalization, and was constructed by Bento Soares and
                     M.Fatima Bonaldo.  "
BASE COUNT      121 a     66 c     74 g    162 t
ORIGIN


  Query Match             1.1%;  Score 38.8;  DB 26;  Length 423;
  Best Local Similarity   51.8%;  Pred. No. 3.9;
  Matches   88;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy   2673 TCTATACCGATTTTTATCAGGCCTTCAGTGATATTCTTTCCAAAATGGGAGGCTGGTTAT 2732
            |||  |   |  ||||  |     ||| | |   |      ||||| | |  ||  ||
Db    257 TCTGAAAATACATTTAAAAAAAACTCATTTAGTGTTAGGGTAAAATAGAACCCTTATTTG 198

Qy   2733 TACCAGGTAAGGACGGTAATACCGTTAAGCTAGATGTTACCTCACTCAAAAATGATTTAA 2792
            ||  |   | |  || |  | |||  ||||  |||   ||  ||||  ||||| | |||||
Db    197 TAACCATTCATTACAGACAAACCAATAAGGAAGAAACTAGGTCACCTAAAAAAGTTTTAA 138

Qy   2793 ACAGTTTAGTCAATAAATATAATCAAATAAACAGTAATACCGTTTTATTT 2842
            ||||||||    |   | |  |    | | |   | ||  |||| |||
Db    137 ACAGTTTATATGAATGAGAACACATGCTTTAAACACAAACTATTTTTTTT 88


RESULT  14
W57781
LOCUS       W57781         325 bp    mRNA            EST       15-OCT-1996
DEFINITION  zd17b03.r1 Soares_fetal_heart_NbHH19W Homo sapiens cDNA clone
            IMAGE:340877 5' similar to gb:X52104 P68 PROTEIN (HUMAN);, mRNA
            sequence.
ACCESSION   W57781
NID         g1364545
VERSION     W57781.1  GI:1364545
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 325)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     On Apr 5, 1995 this sequence version replaced gi:760479.

            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1879   Std Error: 0.00
            Seq primer: mob.REGA+ET
            High quality sequence stop: 247.
FEATURES             Location/Qualifiers
     source          1. .325
                     /organism="Homo sapiens"
                     /db_xref="GDB:1266252"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:340877"
                     /clone_lib="Soares_fetal_heart_NbHH19W"
                     /sex="unknown"
                     /dev_stage="19 weeks"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: heart; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATCTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by
                     M.Fatima Bonaldo. This library was constructed from the
                     same fetus as the fetal lung library, Soares fetal lung
                     NbHL19W."
BASE COUNT      113 a     50 c     75 g     83 t      4 others
ORIGIN


  Query Match             1.1%;  Score 38.8;  DB 26;  Length 325;
  Best Local Similarity   49.7%;  Pred. No. 3.5;
  Matches   97;  Conservative    0;  Mismatches   98;  Indels    0;  Gaps    0;

Qy   2915 TTACCGAATAGCTGCCTGAAATCTTATGGATCCGGTTATGTCGTCACCGTTGATCTGACG 2974
            || | ||     | |||||||   |||          |  | |   |  |||| ||||
Db    120 TTGCTGAGAGATTTCCTGAAAGACTATATTCATATAAACATTGGTGCACTTGAACTGAGT 179

Qy   2975 CCATTACAAAAAATGGTTCAGGATATTGATGGTTTAGGCGCGCCGGGAAAAGACTCAAAA 3034
            ||   || || ||  |||||  | |||   || |        |   ||||||  | ||||
Db    180 GCAAACCACAACATTCTTCAGATTGTTGGATGTGTGTCATGACGTAGAAAAGGATGAAAA 239

Qy   3035 CTCGAAATGGATAACGCCAAATATCAAGCCTGGCAGTCGGGTTTTAAAGCGCAGGAAGAA 3094
            |   |  ||| |  | |||||| |  ||  |  || | || ||| |   |
Db    240 ACTTATTCGTCTAATGGAAGATATCATGAGTGAGAAACGAGATTAAAACCATTTGNTTTT 299

Qy   3095 AATATGAAAACCACA 3109
          | || |||| | | |
Db    300 AGTAAGAAACCAAAA 314


RESULT  15
AA174511/c
LOCUS       AA174511       513 bp    mRNA            EST       16-FEB-1997
DEFINITION  ms69h04.r1 Soares mouse 3NbMS Mus musculus cDNA clone IMAGE:616855
            5', mRNA sequence.
ACCESSION   AA174511
NID         g1755634
VERSION     AA174511.1  GI:1755634
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 513)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
```

```
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Apr 14, 1993 this sequence version replaced gi:692920.

            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:377679
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 493.
FEATURES             Location/Qualifiers
     source          1. .513
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /map="1"
                     /clone="IMAGE:616855"
                     /clone_lib="Soares mouse 3NbMS"
                     /sex="male"
                     /tissue_type="Spleen"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGCTGTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. RNA
                     provided by Dr. Bertrand Jordan. Library went through
                     three rounds of normalization, and was constructed by
                     Bento Soares and M.Fatima Bonaldo."
BASE COUNT      139 a     85 c    100 g    189 t
ORIGIN


  Query Match             1.1%;  Score 38.8;  DB 29;  Length 513;
  Best Local Similarity   51.8%;  Pred. No. 4.2;
  Matches   88;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy   2673 TCTATACCGATTTTTATCAGGCCTTCAGTGATATTCTTTCCAAAATGGGAGGCTGGTTAT 2732
            |||  |   |  ||||  |     ||| | |   |      ||||| | |  ||  ||
Db    290 TCTGAAAATACATTTAAAAAAAACTCATTTAGTGTTAGGGTAAAATAGAACCCTTATTTG 231

Qy   2733 TACCAGGTAAGGACGGTAATACCGTTAAGCTAGATGTTACCTCACTCAAAAATGATTTAA 2792
           || |   | |  || | | ||| ||||  |||   ||  ||||  ||||| | |||||
Db    230 TAACCATTCATTACAGACAAACCAATAAGGAAGAAACTAGGTCACCTAAAAAAGTTTTAA 171

Qy   2793 ACAGTTTAGTCAATAAATATAATCAAATAAACAGTAATACCGTTTTATTT 2842
           ||||||||    |   | | |     |  | |   | ||  |||| |||
Db    170 ACAGTTTATATGAATGAGAACACATGCTTTAAACACAAACTATTTTTTTT 121


Search completed: October  5, 1999, 17:14:21
Job time: 8704 sec
```

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 1999, 18:01:53 ; Search time 140.41 Seconds
(without alignments)
2549.771 Million cell updates/sec

Title: US-09-068-804-15
Perfect score: 3441
Sequence: 1 CGCAAAGCCGAGGAAACGAA.....AAACAGGCGAGTCTGGATGC 3441

Scoring table: IDENTITY_NUC

Searched: 192659 seqs, 52021692 residues

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/5C_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/5D_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	56.6	1.6	7218	2	US-08-232-463-14	Sequence 14, Appl
	2	45.2	1.3	1505	1	US-07-915-246-1	Sequence 1, Appli
c	3	38.2	1.1	10366	2	US-08-246-982A-5	Sequence 5, Appli
c	4	38.2	1.1	10366	2	US-08-453-265-5	Sequence 5, Appli
c	5	38.2	1.1	10348	3	US-08-457-273B-41	Sequence 41, Appl
c	6	35.8	1.0	1295	2	US-08-433-854-3	Sequence 3, Appli
	7	35.8	1.0	2793	2	US-08-209-747-1	Sequence 1, Appli
c	8	35.8	1.0	1295	2	US-08-174-745A-3	Sequence 3, Appli
	9	35.8	1.0	2793	2	US-08-458-298-1	Sequence 1, Appli
c	10	35.8	1.0	1295	3	US-08-195-947-3	Sequence 3, Appli
c	11	35.8	1.0	1295	3	US-08-433-885-3	Sequence 3, Appli
	12	35	1.0	4940	1	US-08-484-105-1	Sequence 1, Appli
	13	35	1.0	4940	1	US-08-484-106-1	Sequence 1, Appli
	14	34.8	1.0	6530	4	US-08-146-930-1	Sequence 1, Appli
	15	34.8	1.0	6530	5	PCT-US93-03993-1	Sequence 1, Appli
	16	34.4	1.0	5738	1	US-08-409-995-3	Sequence 3, Appli
	17	34.2	1.0	10747	4	US-08-147-777-1	Sequence 1, Appli
	18	34.2	1.0	24979	4	US-08-147-777-3	Sequence 3, Appli
	19	34.2	1.0	10747	5	PCT-US93-03985-1	Sequence 1, Appli
	20	34.2	1.0	24979	5	PCT-US93-03985-3	Sequence 3, Appli
	21	34	1.0	2338	2	US-08-425-069-1	Sequence 1, Appli
	22	33.6	1.0	2784	2	US-08-471-454-1	Sequence 1, Appli
	23	33.6	1.0	2784	3	US-08-466-974-1	Sequence 1, Appli
	24	33.6	1.0	2784	4	US-08-471-453-1	Sequence 1, Appli
	25	33.6	1.0	1755	4	US-08-307-588-3	Sequence 3, Appli
	26	33.4	1.0	3319	1	US-08-006-676B-2	Sequence 2, Appli
	27	33.4	1.0	3319	2	US-08-282-845-1	Sequence 1, Appli
	28	33.4	1.0	8133	2	US-08-480-604A-5	Sequence 5, Appli
	29	33.4	1.0	3319	4	US-08-428-414A-4	Sequence 4, Appli
	30	33.4	1.0	8133	4	US-08-405-496A-5	Sequence 5, Appli
	31	33.4	1.0	3319	5	PCT-US94-00324-2	Sequence 2, Appli
	32	32.8	1.0	10592	3	US-08-680-726A-51	Sequence 51, Appl
c	33	32.8	1.0	10592	3	US-08-680-726A-52	Sequence 52, Appl
	34	32.4	0.9	564	3	US-08-775-978-2	Sequence 2, Appli
	35	32.4	0.9	459	4	US-08-387-942C-9	Sequence 9, Appli
c	36	32.2	0.9	9757	1	US-08-093-453B-1	Sequence 1, Appli
c	37	32.2	0.9	9759	2	US-08-459-041A-1	Sequence 1, Appli
	38	31.8	0.9	894	3	US-08-454-720A-32	Sequence 32, Appl
	39	31.6	0.9	5437	1	US-07-661-610C-1	Sequence 1, Appli
	40	31.6	0.9	624	1	US-07-661-610C-14	Sequence 14, Appl
	41	31.6	0.9	2814	1	US-07-781-034-1	Sequence 1, Appli
	42	31.6	0.9	1288	2	US-08-440-856A-9	Sequence 9, Appli
	43	31.6	0.9	4835	2	US-08-386-495-9	Sequence 9, Appli
	44	31.6	0.9	2814	5	PCT-US92-08328-1	Sequence 1, Appli
	45	31.6	0.9	4255	5	PCT-US96-02331-14	Sequence 14, Appl

ALIGNMENTS

```
RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14


  Query Match             1.6%;  Score 56.6;  DB 2;  Length 7218;
  Best Local Similarity   5.0%;  Pred. No. 2.8e-07;
  Matches   20;  Conservative  221;  Mismatches  160;  Indels    0;  Gaps    0;

Qy    2985 AAATGGTTCAGGATATTGATGGTTTAGGCGCGCCGGGAAAAGACTCAAAACTCGAAATGG 3044
           || | ||| | || || || | || ||  |:  :::::::::   ::::  :::: ::
Db    1465 AAGTAGTTAAAGAGATAGAAGAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1406

Qy    3045 ATAACGCCAAATATCAAGCCTGGCAGTCGGGTTTTAAAGCGCAGGAAGAAAATATGAAAA 3104
           : :: :  ::: :  :::   :: ::  :::    :::: : :::::::::: : :::::
```

```
Db   1405 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1346

Qy   3105 CCACATTACAGACGCTGACGCAAAAATATAGCAATGCCAATTCATTGTACGACAACCTGG 3164
          : :  : ::: :  :: : ::::: : :: :: :  ::   :  : : :: ::   ::
Db   1345 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1286

Qy   3165 TAAAAGTGCTGAGCAGTACGATAAGTAGCAGCCTGGAAACCGCCAAAAGCTTCCTGCAAG 3224
          ::::: :  ::: :: : :: ::: :: ::   :::::  :  :::::      : :::
Db   1285 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1226

Qy   3225 GATAACAGAAGAGGATATTAATAATGGTTACAAGTGTAAGGACTCAGCCCCCCGTCATAA 3284
          :: :: ::::::::: :  :: :: ::  : ::: : :::::   ::      :  : ::
Db   1225 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1166

Qy   3285 TGCCAGGTATGCAGACCGAGATCAAAACGCAGGCCACGAATCTTGCGGCGAATCTTTCCG 3344
          :  ::: : : :::  ::::  :::: : :::  : :::    : :: :::       :
Db   1165 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1106

Qy   3345 CAGTCAGAGAAAGTGCCACAGCGACGCTGTCAGGGGAAATT 3385
          ::  :::::::: :  : :: :: :  :  :::::::: |
Db   1105 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRAT 1065
```

RESULT 2
US-07-915-246-1
; Sequence 1, Application US/07915246
; Patent No. 5401836
; GENERAL INFORMATION:
; APPLICANT: Baszczynski, Chris L.
; APPLICANT: Fallis, Lynne
; APPLICANT: Bellmare, Guy
; APPLICANT: Boivin, Rodolphe
; TITLE OF INVENTION: A BRASSICA REGULATORY SEQUENCE FOR
; TITLE OF INVENTION: ROOT-SPECIFIC OR ROOT-ABUNDANT GENE EXPRESSION
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McAndrews, Held, and Malloy
; STREET: 500 W. Madison St. 34th Floor
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60661
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/915,246
; FILING DATE: 19920716
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pochopien, Donald J.
; REGISTRATION NUMBER: 32,167
; REFERENCE/DOCKET NUMBER: 91 P 1125
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312 707-8889
; TELEFAX: 312 707-9155
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1505 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Brassica napus
; STRAIN: Westar
; DEVELOPMENTAL STAGE: Somatic
; TISSUE TYPE: Root
US-07-915-246-1

Query Match 1.3%; Score 45.2; DB 1; Length 1505;
Best Local Similarity 47.2%; Pred. No. 0.00038;
Matches 137; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

```
Qy    111 GGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGAC 170
          || | |||  ||||||| | | | ||   ||| ||  |  ||  |    | || || |
Db   1047 GGTGGTGGAGGCGGTGGAAATGGAGGCGGTGGAGGAGGTGGAGGTGCACACGGTGGTGGA 1106

Qy    171 GGGAGTGTCGTTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTT 230
            |||  |   |    || || |||  |   |  |||||||   | | || ||  |
Db   1107 TACGGTGGTGGAGAAGGCGCTGGTGCTGGAGGAGGATATGGAGGTGGCGGTGCAGGTGGA 1166

Qy    231 AATGGAGCTGATTGGCAAGGCGATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAAGAA 290
            |||||| || | |    || |   |   | ||| |||| |    |   | | |
Db   1167 CATGGAGGTGGTGGAGGCGGGGGAAAAGGAGGCGGTGGAGGAGGAGGTTCTGGCGCCGGT 1226

Qy    291 AACGGCAGAGATGGCCGGCAGCATTGTTGGTGCGATTGTCGCCGCTATTGCCATGGTGGC 350
            |    | |||  |  |   |   |  | |  |   | |       | |  ||||||
Db   1227 GGAGCTCACGGTGGTGGTTATGGTGCCGGAGGTGGAGCTGGAGAGGGATACGGTGGTGGT 1286

Qy    351 GGTCATTGTGGTGGTCGCAGTTGTCGGGAAAGGCGCGGCGGCGAAACTGG 400
          |||      || || | | || || | | | |  | || || | | |||||
Db   1287 GGTGGAGAAGGAGGACACGGTGGTGGAGGAGGCGGTGGTGGTGGAGCTGG 1336
```

RESULT 3
US-08-246-982A-5/c
; Sequence 5, Application US/08246982A
; Patent No. 5686288
; GENERAL INFORMATION:
; APPLICANT: MacDonald, Marcy E.
; APPLICANT: Ambrose, Christine M.
; APPLICANT: Duyao, Mabel P.
; APPLICANT: Gusella, James F.
; TITLE OF INVENTION: Huntingtin DNA, Protein And Uses Thereof
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/246,982A
; FILING DATE: May 20, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldstein, Jorge, A.
; REGISTRATION NUMBER: 29,021
; REFERENCE/DOCKET NUMBER: 0609.3880002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10366 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 316..9748
US-08-246-982A-5

```
  Query Match             1.1%;  Score 38.2;  DB 2;  Length 10366;
  Best Local Similarity   51.6%;  Pred. No. 0.26;
  Matches  111;  Conservative    0;  Mismatches  103;  Indels    1;  Gaps    1;

Qy      74 GCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTG 133
           ||   ||||| || |      | |  ||||||  |  |||   ||| || | ||   |
Db     552 GCCGGGTGGCGGCGGGGGCGGCGGCGGGGGCGGCTGCGGCTGAGGCAGCAGCGGCTGTGC 493

Qy     134 CGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCGTTTAT-TCAGCAG 192
           | |    || ||| || |||  || | ||||||| ||| | | ||    |  | || |
Db     492 CTGCGGCGGCGGCTGAGGAAGCTGAGGAGGCGGCGGCGGCGGCGGCGGTGGCGGCTG 433

Qy     193 GCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTGATTGGCAAGGCG 252
           |||    | | |  ||  ||   |||||  || | |  ||  |||| |      |  |
Db     432 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 373

Qy     253 ATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAA 287
           |     |||  || || ||| | |  | |  | |
Db     372 CTGCTGGAAGGACTTGAGGGACTCGAAGGCCTTCA 338


RESULT   4
US-08-453-265-5/c
; Sequence 5, Application US/08453265
; Patent No. 5693757
;  GENERAL INFORMATION:
;    APPLICANT:  MacDonald, Marcy E.
;    APPLICANT:  Ambrose, Christine M.
;    APPLICANT:  Duyao, Mabel P.
;    APPLICANT:  Gusella, James F.
;    TITLE OF INVENTION:  Huntingtin DNA, Protein And Uses Thereof
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox
;      STREET:  1100 New York Avenue
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/453,265
;      FILING DATE:  30-MAY-1995
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ludwig, Steven R.
;      REGISTRATION NUMBER:  36,203
;      REFERENCE/DOCKET NUMBER:  0609.3880003
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10366 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  316..9748
US-08-453-265-5


  Query Match             1.1%;  Score 38.2;  DB 2;  Length 10366;
  Best Local Similarity   51.6%;  Pred. No. 0.26;
  Matches  111;  Conservative    0;  Mismatches  103;  Indels    1;  Gaps    1;

Qy      74 GCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTG 133
           ||   ||||| || |      | |  ||||||  |  |||   ||| || | ||   |
Db     552 GCCGGGTGGCGGCGGGGGCGGCGGCGGGGGCGGCTGCGGCTGAGGCAGCAGCGGCTGTGC 493

Qy     134 CGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCGTTTAT-TCAGCAG 192
           | |    || ||| || |||  || | ||||||| ||| | | ||    |  | || |
Db     492 CTGCGGCGGCGGCTGAGGAAGCTGAGGAGGCGGCGGCGGCGGCGGCGGTGGCGGCTG 433

Qy     193 GCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTGATTGGCAAGGCG 252
           |||    | | |  ||  ||   |||||  || | |  ||  |||| |      |  |
Db     432 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 373

Qy     253 ATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAA 287
           |     |||  || || ||| | |  | |  | |
Db     372 CTGCTGGAAGGACTTGAGGGACTCGAAGGCCTTCA 338

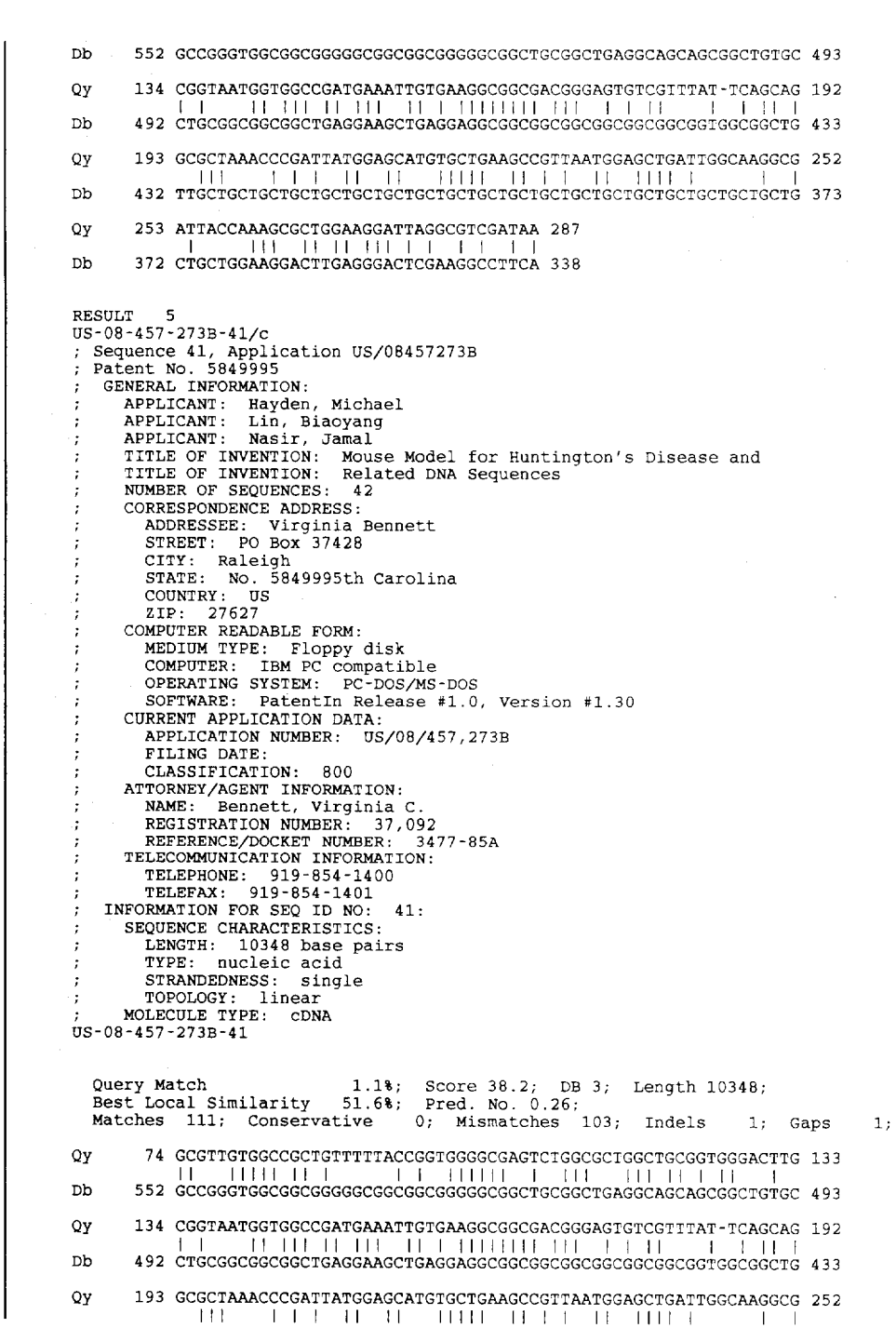

RESULT   5
US-08-457-273B-41/c
; Sequence 41, Application US/08457273B
; Patent No. 5849995
;  GENERAL INFORMATION:
;    APPLICANT:  Hayden, Michael
;    APPLICANT:  Lin, Biaoyang
;    APPLICANT:  Nasir, Jamal
;    TITLE OF INVENTION:  Mouse Model for Huntington's Disease and
;    TITLE OF INVENTION:  Related DNA Sequences
;    NUMBER OF SEQUENCES:  42
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Virginia Bennett
;      STREET:  PO Box 37428
;      CITY:  Raleigh
;      STATE:  No. 5849995th Carolina
;      COUNTRY:  US
;      ZIP:  27627
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/457,273B
;      FILING DATE:
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bennett, Virginia C.
;      REGISTRATION NUMBER:  37,092
;      REFERENCE/DOCKET NUMBER:  3477-85A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-854-1400
;      TELEFAX:  919-854-1401
;  INFORMATION FOR SEQ ID NO:  41:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10348 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-457-273B-41


  Query Match             1.1%;  Score 38.2;  DB 3;  Length 10348;
  Best Local Similarity   51.6%;  Pred. No. 0.26;
  Matches  111;  Conservative    0;  Mismatches  103;  Indels    1;  Gaps    1;

Qy      74 GCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTG 133
           ||   ||||| || |      | |  ||||||  |  |||   ||| || | ||   |
Db     552 GCCGGGTGGCGGCGGGGGCGGCGGCGGGGGCGGCTGCGGCTGAGGCAGCAGCGGCTGTGC 493

Qy     134 CGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCGTTTAT-TCAGCAG 192
           | |    || ||| || |||  || | ||||||| ||| | | ||    |  | || |
Db     492 CTGCGGCGGCGGCTGAGGAAGCTGAGGAGGCGGCGGCGGCGGCGGCGGTGGCGGCTG 433

Qy     193 GCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTGATTGGCAAGGCG 252
           |||    | | |  ||  ||   |||||  || | |  ||  |||| |      |  |
```



```
Db    432 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 373

Qy    253 ATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAA 287
          |   |||  || || ||| | |  | |  | |
Db    372 CTGCTGGAAGGACTTGAGGGACTCGAAGGCCTTCA 338


RESULT   6
US-08-433-854-3/c
; Sequence 3, Application US/08433854
; Patent No. 5721119
;  GENERAL INFORMATION:
;    APPLICANT:  Singh, Mohan Bir
;    APPLICANT:  Knox, Robert B.
;    APPLICANT:  Smith, Penelope
;    APPLICANT:  Avjioglu, Asil
;    APPLICANT:  Theerakulpisut, Piyada
;    APPLICANT:  Hough, Terryn
;    APPLICANT:  Suphioglu, Cenk
;    APPLICANT:  Ong, Eng Kok
;    TITLE OF INVENTION:  Ryegrass Pollen Allergen
;    NUMBER OF SEQUENCES:  53
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock Washburn Kurtz Mackiewicz & No. 5721119ris
;      STREET:  One Liberty Place - 46th Floor
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/433,854
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/930,060
;      FILING DATE:  14-AUG-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hohenschutz, Liza D.
;      REGISTRATION NUMBER:  33,712
;      REFERENCE/DOCKET NUMBER:  IMPH-0024
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 568-3100
;      TELEFAX:  (215) 568-3439
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1295 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  25..1041
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  100..1041
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  25..99
US-08-433-854-3


  Query Match             1.0%;  Score 35.8;  DB 2;  Length 1295;
  Best Local Similarity   59.2%;  Pred. No. 0.33;
  Matches   61;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy     95 CCGGTGGGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTGGCCGATGAAA 154
          | |  | |||| |  | ||| |||| |||| |||  |||||||   |||||| |  |
```

```
Db   1027 CAGCAGCGGCGGGGGTTGCGGTGGCAGCGGCGGGGGTTGCGGTGGCGGTGGCAGCGGCGG 968

Qy    155 TTGTGAAGGCGGCGACGGGAGTGTCGTTTATTCAGCAGGCGCT 197
            ||   || ||||  || || | ||    ||  |  |||| |
Db    967 GGGTAGCGGTGGCGTAGGCAGCGGCGGGGGTTGCGGTGGCGGT 925


RESULT   7
US-08-209-747-1
; Sequence 1, Application US/08209747
; Patent No. 5733771
;  GENERAL INFORMATION:
;    APPLICANT:  Lewis, Randolph V.
;    APPLICANT:  Colgin, Mark
;    TITLE OF INVENTION:  cDNAs Encoding Minor Ampullate Spider
;    TITLE OF INVENTION:  Silk Proteins
;    NUMBER OF SEQUENCES:  56
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Birch, Stewart, Kolasch & Birch
;      STREET:  P.O. Box 747
;      CITY:  Falls Church
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22040-3487
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/209,747
;      FILING DATE:  14-MAR-1994
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Murphy Jr., Gerald M.
;      REGISTRATION NUMBER:  28,977
;      REFERENCE/DOCKET NUMBER:  1447-104P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-205-8000
;      TELEFAX:  703-205-8050
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2793 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Nephila clavipes
;      TISSUE TYPE:  minor ampullate gland
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  183..2675
;      OTHER INFORMATION:  /product= "N. clavipes minor
;      OTHER INFORMATION:  ampullate silk protein"
US-08-209-747-1


  Query Match             1.0%;  Score 35.8;  DB 2;  Length 2793;
  Best Local Similarity   55.1%;  Pred. No. 0.58;
  Matches   70;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy    109 CTGGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCG 168
          | || |||| | ||| |||   || ||| |  ||||   || |  |  ||| |   || |
Db    454 CAGGAGCTGCCGGCGCTGGAGGTTACGGAAGAGGTGCTGGAGGCTACGGTGGACAAGGAG 513

Qy    169 ACGGGAGTGTCGTTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCG 228
              ||| ||       ||| || |||   | | |  ||||||| | || | || ||
Db    514 GATATGGTGCCGGAGCCGGAGCTGGTGCTGCTGCAGCTGCTGGAGCAGGAGCCGGAGGCG 573

Qy    229 TTAATGG 235
            |  |||
```

```
Db    574 CTGGTGG 580


RESULT   8
US-08-174-745A-3/c
; Sequence 3, Application US/08174745A
; Patent No. 5736362
;  GENERAL INFORMATION:
;    APPLICANT:  Singh, Mohan Bir
;    APPLICANT:  Knox, Robert B.
;    APPLICANT:  Smith, Penelope
;    APPLICANT:  Avjioglu, Asil
;    APPLICANT:  Theerakulpisut, Piyada
;    APPLICANT:  Hough, Terryn
;    APPLICANT:  Suphioglu, Cenk
;    APPLICANT:  Ong, Eng Kok
;    TITLE OF INVENTION:  Ryegrass Pollen Allergen
;    NUMBER OF SEQUENCES:  53
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock Washburn Kurtz Mackiewicz & No. 5736362ris
;      STREET:  One Liberty Place - 46th Floor
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/174,745A
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/930,060
;      FILING DATE:  14-AUG-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hohenschutz, Liza D.
;      REGISTRATION NUMBER:  33,712
;      REFERENCE/DOCKET NUMBER:  IMPH-0024
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 568-3100
;      TELEFAX:  (215) 568-3439
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1295 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  25..1041
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  100..1041
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  25..99
US-08-174-745A-3


  Query Match             1.0%;  Score 35.8;  DB 2;  Length 1295;
  Best Local Similarity   59.2%;  Pred. No. 0.33;
  Matches   61;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy      95 CCGGTGGGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTGGCCGATGAAA 154
           | |  | |||| |  | ||| |||| |||| |||  |||||||   ||||||  |  |
Db    1027 CAGCAGCGGCGGGGGTTGCGGTGGCAGCGGCGGGGGTTGCGGTGGCGGTGGCAGCGGCGG 968

Qy     155 TTGTGAAGGCGGCGACGGGAGTGTCGTTTATTCAGCAGGCGCT 197
             ||   || ||||  || || | ||    ||  |  |||| |


Db    967 GGGTAGCGGTGGCGTAGGCAGCGGCGGGGGTTGCGGTGGCGGT 925


RESULT   9
US-08-458-298-1
; Sequence 1, Application US/08458298
; Patent No. 5756677
;  GENERAL INFORMATION:
;    APPLICANT:  Lewis, Randolph V.
;    APPLICANT:  Colgin, Mark
;    TITLE OF INVENTION:  cDNAs Encoding Minor Ampullate Spider
;    TITLE OF INVENTION:  Silk Proteins
;    NUMBER OF SEQUENCES:  56
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Birch, Stewart, Kolasch & Birch
;      STREET:  P.O. Box 747
;      CITY:  Falls Church
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22040-3487
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/458,298
;      FILING DATE:  02-JUN-1995
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/209,747
;      FILING DATE:  14-MAR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Murphy Jr., Gerald M.
;      REGISTRATION NUMBER:  28,977
;      REFERENCE/DOCKET NUMBER:  1447-104P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-205-8000
;      TELEFAX:  703-205-8050
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2793 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Nephila clavipes
;      TISSUE TYPE:  minor ampullate gland
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  183..2675
;      OTHER INFORMATION:  /product= "N. clavipes minor
;      OTHER INFORMATION:  ampullate silk protein"
US-08-458-298-1


  Query Match             1.0%;  Score 35.8;  DB 2;  Length 2793;
  Best Local Similarity   55.1%;  Pred. No. 0.58;
  Matches   70;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy     109 CTGGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCG 168
           | || |||| | ||| |||   || ||| |  ||||   || |  |  ||| |   || |
Db     454 CAGGAGCTGCCGGCGCTGGAGGTTACGGAAGAGGTGCTGGAGGCTACGGTGGACAAGGAG 513

Qy     169 ACGGGAGTGTCGTTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCG 228
                 ||| ||        ||| || |||    | | |  ||||||| | || | || ||
Db     514 GATATGGTGCCGGAGCCGGAGCTGGTGCTGCTGCAGCTGCTGGAGCAGGAGCCGGAGGCG 573

Qy     229 TTAATGG 235
             |  |||
Db     574 CTGGTGG 580
```

```
RESULT 10
US-08-195-947-3/c
; Sequence 3, Application US/08195947
; Patent No. 5840316
;  GENERAL INFORMATION:
;    APPLICANT:  Singh, Mohan Bir
;    APPLICANT:  Knox, Robert B.
;    APPLICANT:  Smith, Penelope
;    APPLICANT:  Avjioglu, Asil
;    APPLICANT:  Theerakulpisut, Piyada
;    APPLICANT:  Hough, Terryn
;    APPLICANT:  Suphioglu, Cenk
;    APPLICANT:  Ong, Eng Kok
;    TITLE OF INVENTION:  Ryegrass Pollen Allergen
;    NUMBER OF SEQUENCES:  53
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock Washburn Kurtz Mackiewicz & No. 5840316ris
;      STREET:  One Liberty Place - 46th Floor
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/195,947
;      FILING DATE:  14-FEB-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/930,060
;      FILING DATE:  14-AUG-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hohenschutz, Liza D.
;      REGISTRATION NUMBER:  33,712
;      REFERENCE/DOCKET NUMBER:  IMPH-0024
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 568-3100
;      TELEFAX:  (215) 568-3439
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1295 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  25..1041
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  100..1041
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  25..99
US-08-195-947-3


  Query Match             1.0%;  Score 35.8;  DB 3;  Length 1295;
  Best Local Similarity   59.2%;  Pred. No. 0.33;
  Matches   61;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy      95 CCGGTGGGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTGGCCGATGAAA 154
           | |  | ||||  |  | ||| |||| |||| |||  |||||||   ||||||  |  |
Db    1027 CAGCAGCGGCGGGGGTTGCGGTGGCAGCGGCGGGGGTTGCGGTGGCGGTGGCAGCGGCGG 968

Qy     155 TTGTGAAGGCGGCGACGGGAGTGTCGTTTATTCAGCAGGCGCT 197
             ||   || ||||  || || | ||    ||  |  ||||
Db     967 GGGTAGCGGTGGCGTAGGCAGCGGCGGGGGTTGCGGTGGCGGT 925
```

```
RESULT 11
US-08-433-885-3/c
; Sequence 3, Application US/08433885
; Patent No. 5869333
;  GENERAL INFORMATION:
;    APPLICANT:  Singh, Mohan Bir
;    APPLICANT:  Knox, Robert B.
;    APPLICANT:  Smith, Penelope
;    APPLICANT:  Avjioglu, Asil
;    APPLICANT:  Theerakulpisut, Piyada
;    APPLICANT:  Hough, Terryn
;    APPLICANT:  Suphioglu, Cenk
;    APPLICANT:  Ong, Eng Kok
;    TITLE OF INVENTION:  Ryegrass Pollen Allergen
;    NUMBER OF SEQUENCES:  53
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lahive & Cockfield
;      STREET:  60 State Street, Suite 510
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/433,885
;      FILING DATE:  02-MAY-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mandragouras, Amy E.
;      REGISTRATION NUMBER:  36,207
;      REFERENCE/DOCKET NUMBER:  IMI-039C2D2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 227-7400
;      TELEFAX:  (617) 227-5941
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1295 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  25..1041
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  100..1041
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  25..99
US-08-433-885-3


  Query Match             1.0%;  Score 35.8;  DB 3;  Length 1295;
  Best Local Similarity   59.2%;  Pred. No. 0.33;
  Matches   61;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy      95 CCGGTGGGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTGGCCGATGAAA 154
           | |  | ||||  |  | ||| |||| |||| |||  |||||||   ||||||  |  |
Db    1027 CAGCAGCGGCGGGGGTTGCGGTGGCAGCGGCGGGGGTTGCGGTGGCGGTGGCAGCGGCGG 968

Qy     155 TTGTGAAGGCGGCGACGGGAGTGTCGTTTATTCAGCAGGCGCT 197
             ||   || ||||  || || | ||    ||  |  ||||
Db     967 GGGTAGCGGTGGCGTAGGCAGCGGCGGGGGTTGCGGTGGCGGT 925


RESULT 12
```

```
US-08-484-105-1
; Sequence 1, Application US/08484105
; Patent No. 5589341
;  GENERAL INFORMATION:
;    APPLICANT:  STILLMAN, Bruce
;    APPLICANT:  BELL, Stephen P
;    APPLICANT:  KOBAYASHI, Ryuji
;    APPLICANT:  RINE, Jasper
;    APPLICANT:  FOSS, Margit
;    APPLICANT:  McNALLY, Francis J
;    APPLICANT:  LAURENSON, Patricia
;    APPLICANT:  HERSKOWITZ, Ira
;    APPLICANT:  LI, Joachim J
;    APPLICANT:  GAVIN, Kimberly
;    TITLE OF INVENTION:  ORIGIN OF REPLICATION COMPLEX GENES
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/484,105
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman Ph.D., Richard Aron
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  A-59032/DJB/RAO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 494-8700
;      TELEFAX:  (415) 494-8771
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4940 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-484-105-1


  Query Match             1.0%;  Score 35;  DB 1;  Length 4940;
  Best Local Similarity   54.2%;  Pred. No. 1.6;
  Matches   71;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy   1636 AATATGAAGAAAACCGGTACCGATGCGACGAAAAATCTTAATGATGCGACGCTTAAATCT 1695
          ||   ||  ||||  ||| |  |   ||   | ||   || | |  || | |  |  ||
Db   1334 AAAGAGATCAAAAGAGGTCCTCAAAAGAAAGATAAGGCTACTCAAACGGCACAAATTTCA 1393

Qy   1696 AATGCCGGAACCAGCGCCACGGAAAGTCTGGGTATTAAAGACAGTAATAAACAAATCTCC 1755
          | || | ||| || || || || |    || || | | ||| ||||| || |    ||
Db   1394 GACGCAGAAACAAGAGCTACAGATATAACGGATAATGAGGACGGTAATGAAGATGAATCA 1453

Qy   1756 CCTGAACATCA 1766
          ||||  || |
Db   1454 TCTGATTATGA 1464


RESULT  13
US-08-484-106-1
; Sequence 1, Application US/08484106
; Patent No. 5614618
;  GENERAL INFORMATION:
;    APPLICANT:  STILLMAN, Bruce
;    APPLICANT:  BELL, Stephen P
;    APPLICANT:  KOBAYASHI, Ryuji
;    APPLICANT:  RINE, Jasper
;    APPLICANT:  FOSS, Margit
;    APPLICANT:  McNALLY, Francis J
;    APPLICANT:  LAURENSON, Patricia
;    APPLICANT:  HERSKOWITZ, Ira
;    APPLICANT:  LI, Joachim J
;    APPLICANT:  GAVIN, Kimberly
;    TITLE OF INVENTION:  ORIGIN OF REPLICATION COMPLEX GENES
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/484,106
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman Ph.D., Richard Aron
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  A-59032/DJB/RAO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 494-8700
;      TELEFAX:  (415) 494-8771
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4940 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-484-106-1


  Query Match             1.0%;  Score 35;  DB 1;  Length 4940;
  Best Local Similarity   54.2%;  Pred. No. 1.6;
  Matches   71;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy   1636 AATATGAAGAAAACCGGTACCGATGCGACGAAAAATCTTAATGATGCGACGCTTAAATCT 1695
          ||   ||  ||||  ||| |  |   ||   | ||   || | |  || | |  |  ||
Db   1334 AAAGAGATCAAAAGAGGTCCTCAAAAGAAAGATAAGGCTACTCAAACGGCACAAATTTCA 1393

Qy   1696 AATGCCGGAACCAGCGCCACGGAAAGTCTGGGTATTAAAGACAGTAATAAACAAATCTCC 1755
          | || | ||| || || || || |    || || | | ||| ||||| || |    ||
Db   1394 GACGCAGAAACAAGAGCTACAGATATAACGGATAATGAGGACGGTAATGAAGATGAATCA 1453

Qy   1756 CCTGAACATCA 1766
          ||||  || |
Db   1454 TCTGATTATGA 1464


RESULT  14
US-08-146-930-1
; Sequence 1, Application US/08146930
; Patent No. 5958764
;  GENERAL INFORMATION:
;    APPLICANT:  Roop, Dennis R.
;    APPLICANT:  Rothnagel, Joseph A.
;    APPLICANT:  Greenhalgh, David A.
;    TITLE OF INVENTION:  SPECIFIC EXPRESSION VECTORS
;    TITLE OF INVENTION:  AND METHODS OF USE
;    NUMBER OF SEQUENCES:  4
```

```
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LYON & LYON
;      STREET:  611 West Sixth Street
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90017
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb storage
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  IBM MS-DOS (Version 5.0)
;      SOFTWARE:  WordPerfect (Version 5.1)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/146,930
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;    PRIOR APPLICATION DATA:  including application
;    PRIOR APPLICATION DATA:  described below:              two
;      APPLICATION NUMBER:  07/876,286
;      FILING DATE:  April 30, 1992
;      APPLICATION NUMBER:  No. 5958764 yet assigned (204/132)
;      FILING DATE:  October 29, 1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warburg, Richard J.
;      REGISTRATION NUMBER:  32,327
;      REFERENCE/DOCKET NUMBER:  204/152
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6530 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-146-930-1


  Query Match             1.0%;  Score 34.8;  DB 4;  Length 6530;
  Best Local Similarity   44.3%;  Pred. No. 2.2;
  Matches  141;  Conservative    0;  Mismatches  177;  Indels    0;  Gaps    0;

Qy      76 GTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTGCG 135
           || | ||| || | | |  | |  | ||||  ||   ||  ||| |||| ||    | |
Db    3148 GTGGCGGCGGCGGCTCTAGCTGCGGCGGCGGCTCCTCCGGGGGCGGCGGCGGCGGCTCCA 3207

Qy     136 GTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCGTTTATTCAGCAGGCG 195
           |    || ||| ||| |    | |  ||| | || |  | | |   | |  ||  |||
Db    3208 GCTGCGGAGGCGGATCAGGAGGCGGCGGCTCCTACTGCGGAGGCTCCTCTGGAGGCGGCA 3267

Qy     196 CTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTGATTGGCAAGGCGATT 255
                 |  |  || |   || | | || ||  ||     ||| | |    || | |
Db    3268 GCTCCGGTGGCTGCGGCGGCGGTTCCGGAGGCGGCAAGTACTCTGGTGGCGGCGGTGGCT 3327

Qy     256 ACCAAAGCGCTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAGATGGCCGGCAGCATT 315
           |   |||  ||  |   |   | |   | || | || | | |       |
Db    3328 CCAGCTGCGGAGGCGGCTATTCCGGCGGCGGTGGAAGCAGCGGCGGCTCTAGCTGTGGCG 3387

Qy     316 GTTGGTGCGATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTGGTGGTCGCAGTTGTC 375
           | | | |    || || |    | |    | |||||     | ||   |  | || |
Db    3388 GCGGCTACTCAGGTGGCGGTGGATCCAGCTGCGGCGGCGGCGGCGGCTATTCCGGTGGCG 3447

Qy     376 GGGAAAGGCGCGGCGGCG 393
           | |  | | || ||||  |
Db    3448 GCGGCACGAGCTGCGGAG 3465


RESULT  15
PCT-US93-03993-1
; Sequence 1, Application PC/TUS9303993
```

```
;  GENERAL INFORMATION:
;    APPLICANT:  Roop, Dennis R.
;    APPLICANT:  Rothnagel, Joseph A.
;    APPLICANT:  Greenhalgh, David A.
;    TITLE OF INVENTION:  CONSTITUTIVE AND INDUCIBLE EPIDERMAL
;    TITLE OF INVENTION:  VECTOR SYSTEMS
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fulbright & Jaworski
;      STREET:  1301 McKinney, Suite 5100
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  U.S.A.
;      ZIP:  77010-3095
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/03993
;      FILING DATE:  19930428
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Paul, Thomas D.
;      REGISTRATION NUMBER:  32,714
;      REFERENCE/DOCKET NUMBER:  D-5405
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  713/651-5325
;      TELEFAX:  713/651-5246
;      TELEX:  762829
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6530 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
PCT-US93-03993-1


  Query Match             1.0%;  Score 34.8;  DB 5;  Length 6530;
  Best Local Similarity   44.3%;  Pred. No. 2.2;
  Matches  141;  Conservative    0;  Mismatches  177;  Indels    0;  Gaps    0;

Qy      76 GTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTGCG 135
           || | ||| || | | |  | |  | ||||  ||   ||  ||| |||| ||    | |
Db    3148 GTGGCGGCGGCGGCTCTAGCTGCGGCGGCGGCTCCTCCGGGGGCGGCGGCGGCGGCTCCA 3207

Qy     136 GTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCGTTTATTCAGCAGGCG 195
           |    || ||| ||| |    | |  ||| | || |  | | |   | |  ||  |||
Db    3208 GCTGCGGAGGCGGATCAGGAGGCGGCGGCTCCTACTGCGGAGGCTCCTCTGGAGGCGGCA 3267

Qy     196 CTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTGATTGGCAAGGCGATT 255
                 |  |  || |   || | | || ||  ||     ||| | |    || | |
Db    3268 GCTCCGGTGGCTGCGGCGGCGGTTCCGGAGGCGGCAAGTACTCTGGTGGCGGCGGTGGCT 3327

Qy     256 ACCAAAGCGCTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAGATGGCCGGCAGCATT 315
           |   |||  ||  |   |   | |   | || | || | | |       |
Db    3328 CCAGCTGCGGAGGCGGCTATTCCGGCGGCGGTGGAAGCAGCGGCGGCTCTAGCTGTGGCG 3387

Qy     316 GTTGGTGCGATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTGGTGGTCGCAGTTGTC 375
           | | | |    || || |    | |    | |||||     | ||   |  | || |
Db    3388 GCGGCTACTCAGGTGGCGGTGGATCCAGCTGCGGCGGCGGCGGCGGCTATTCCGGTGGCG 3447

Qy     376 GGGAAAGGCGCGGCGGCG 393
           | |  | | || ||||  |
Db    3448 GCGGCACGAGCTGCGGAG 3465
```

Search completed: October 5, 1999, 18:02:19
Job time: 11579 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 1999, 18:00:59 ; Search time 223.92 Seconds
(without alignments)
3844.726 Million cell updates/sec

Title: US-09-068-804-15
Perfect score: 3441
Sequence: 1 CGCAAAGCCGAGGAAACGAA.....AAACAGGCGAGTCTGGATGC 3441

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	3391.8	98.6	3441	1	T67035	Salmonella virulen
	2	1230	35.7	1230	1	T67029	Salmonella secrete
	3	1030.4	29.9	1032	1	T67030	Salmonella secrete
	4	868.4	25.2	870	1	T67028	Salmonella secrete
	5	194	5.6	194	1	T67031	Salmonella secrete
	6	178.6	5.2	1743	1	T85939	Shigella flexneri
c	7	49.4	1.4	1648	1	Q25532	Sequence of genomi
	8	45.2	1.3	1505	1	Q55750	Genomic clone G11F
	9	43.6	1.3	4018	1	Q63879	Polyhydroxyalkanoa
c	10	38.2	1.1	10366	1	Q67401	Huntingtin DNA\IT1
c	11	38.2	1.1	10366	1	T97924	Composite sequence
c	12	38.2	1.1	10366	1	V05828	Human huntingtin c
c	13	38.2	1.1	515	1	X18930	Human huntingtin g
c	14	38	1.1	342	1	Q31795	PRP2 gene. Isolate
c	15	37.8	1.1	5163	1	V20700	Cryptosporidium pa
c	16	37.4	1.1	1107	1	V38816	DNA encoding a dur
	17	37	1.1	864	1	Q03877	Merozoite surface
c	18	36.2	1.1	5318	1	V20701	Cryptosporidium pa
c	19	35.8	1.0	1295	1	Q38230	Sequence of rye gr
	20	35.8	1.0	2744	1	Q98470	MiSP1-containing p
	21	35.8	1.0	1236	1	X15189	DNA encoding a tar
	22	35.6	1.0	1491	1	X15147	Nucleotide sequenc
c	23	35.4	1.0	1523	1	V74861	Staphylococcus aur
	24	35	1.0	2440	1	Q06469	Sequence encoding
	25	35	1.0	3086	1	Q06470	Sequence encoding
	26	35	1.0	4940	1	Q91286	ORC1 gene encoding
	27	35	1.0	4940	1	T73279	S. cerevisiae orig
c	28	34.8	1.0	836	1	Q12218	SMUC-41 intestinal
	29	34.8	1.0	6530	1	Q51557	Loricrin gene. Con
c	30	34.4	1.0	1928	1	Q32139	BsuPI protease inh
	31	34.4	1.0	7291	1	T41476	Haemophilus adhesi
	32	34.4	1.0	799	1	V27329	Streptococcus pneu
	33	34.4	1.0	25002	1	V52181	Streptococcus pneu
c	34	34.4	1.0	13340	1	X23522	O. longistaminata
	35	34.2	1.0	202	1	N81094	Sequence of a synt
	36	34.2	1.0	110000	1	V21209_00	Methanococcus jann
	37	34	1.0	2338	1	Q14183	N.clavipes draglin
	38	34	1.0	459	1	Q63296	Consensus sequence
	39	34	1.0	1173	1	T84527	Plasmid pBNppt ins
	40	34	1.0	2338	1	V23249	Nephila clavipes s
	41	34	1.0	109	1	V16108	Synthetic BRSV Gsy
	42	33.8	1.0	334	1	V88505	EST clone EN108. N
	43	33.8	1.0	114955	1	X53491	Human adenosine A1
c	44	33.8	1.0	114955	1	X53491	Human adenosine A1
	45	33.6	1.0	2784	1	Q11701	Human alpha-interf

ALIGNMENTS

RESULT 1
T67035
ID T67035 standard; DNA; 3441 BP.
AC T67035;
DT 04-AUG-1997 (first entry)
DE Salmonella virulence factor SspBCDA DNA region.
KW Salmonella secreted protein; SspB; SspC; SspD; SspA;
KW bacterial-mediated endocytosis; diagnosis; therapy; vaccine;
KW attenuation; virulence; ss.
OS Salmonella typhimurium.
FH Key Location/Qualifiers
FT cds 1. .870
FT /*tag= a
FT /product= SspB (incomplete)
FT cds 898. .2127
FT /*tag= b
FT /product= SspC
FT cds 2198. .3228
FT /*tag= c
FT /product= SspD
FT cds 3247. .3441
FT /*tag= d
FT /product= SspA (incomplete)
PN WO9718225-A1.
PD 22-MAY-1997.
PF 14-NOV-1996; U18504.
PR 14-NOV-1995; US-006733.
PA (GEHO) GEN HOSPITAL CORP.
PI Miller SI;
DR WPI; 97-289217/26.
DR P-PSDB; W15268-71.
PT New isolated Salmonella secreted proteins and related genes - used
PT to develop products for the detection, treatment or prevention of
PT Salmonella infections
PS Claim 11; Fig 28; 95pp; English.
CC A DNA molecule (T67035) comprises the SspBCDA locus (truncated at
CC the 5' and 3' ends) including the partial sspB, sspC, sspD and
CC partial sspA sequences (see also T67028-31) coding for Salmonella
CC typhimurium secreted proteins SspB, SspC, SspD and SspA (W15268-
CC 70). To identify the ssp genes, S. typhimurium mutants were
CC investigated in invasion assays on HEp-2 epithelial cells. 2
CC Ssp mutants with transposon insertions located between spaT and
CC prgH (see also T67036) were identified. One mutant lacked Ssp of
CC 87, 42 and 36 kDa. The invasiveness of this mutant implicated the
CC 42 and 36 kDa Ssp in invasion. DNA sequencing of this region
CC identified the 4 ssp (partial) genes. The ssp genes can be used in
CC methods for diagnosing Salmonella infection and in methods of
CC targetting an antigen to epithelial cells and of inducing a
CC cytotoxic T cell immune response in a mammal. Ssp gene mutation
CC can be used to create attenuated Salmonella strains useful as live
CC vaccines.
SQ Sequence 3441 BP; 978 A; 761 C; 925 G; 777 T;

Query Match 98.6%; Score 3391.8; DB 1; Length 3441;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 3437; Conservative 0; Mismatches 2; Indels 4; Gaps 4;

```
Qy       1 CGCAAAGCCGAGGAAACGAACCGCATTATGGGATGTATCGGGAAAGTCCTCGGCGCGCTG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CGCAAAGCCGAGGAAACGAACCGCATTATGGGATGTATCGGGAAAGTCCTCGGCGCGCTG 60

Qy      61 CTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCT 120
```


Qy 121 GCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCG 180
Db 121 GCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCG 180
Qy 181 TTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTG 240
Db 181 TTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTG 240
Qy 241 ATTGGCAAGGCGATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAG 300
Db 241 ATTGGCAAGGCGATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAG 300
Qy 301 ATGGCCGGCAGCATTGTTGGTGCGATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTG 360
Db 301 ATGGCCGGCAGCATTGTTGGTGCGATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTG 360
Qy 361 GTGGTCGCAGTTGTCGGGAAAGGCGCGGCGGCGAAACTGGGTAACGCGCTGAGCAAAATG 420
Db 361 GTGGTCGCAGTTGTCGGGAAAGGCGCGGCGGCGAAACTGGGTAACGCGCTGAGCAAAATG 420
Qy 421 ATGGGCGAAACGATTAAGAAGTTGGTGCCTAACGTGCTGAAACAGTTGGCGCAAAACGGC 480
Db 421 ATGGGCGAAACGATTAAGAAGTTGGTGCCTAACGTGCTGAAACAGTTGGCGCAAAACGGC 480
Qy 481 AGCAAACTCTTTACCCAGGGGATGCAACGTATTACTAGCGGTCTGGGTAATGTGGGTAGC 540
Db 481 AGCAAACTCTTTACCCAGGGGATGCAACGTATTACTAGCGGTCTGGGTAATGTGGGTAGC 540
Qy 541 AAGATGGGCCTGCAAACGAATGCCTTAAGTAAAGAGCTGGTAGGTAATACCCTAAATAAA 600
Db 541 AAGATGGGCCTGCAAACGAATGCCTTAAGTAAAGAGCTGGTAGGTAATACCCTAAATAAA 600
Qy 601 GTGGCGTTGGGCATGGAAGTCACGAATACCGCAGCCCAGTCAGCCGGTGGTGTTGCCGAG 660
Db 601 GTGGCGTTGGGCATGGAAGTCACGAATACCGCAGCCCAGTCAGCCGGTGGTGTTGCCGAG 660
Qy 661 GGCGTATTTATTAAAAATGCCAGCGAGGCGCTTGCTGATTTTATGCTCGCCCGTTTTGCC 720
Db 661 GGCGTATTTATTAAAAATGCCAGCGAGGCGCTTGCTGATTTTATGCTCGCCCGTTTTGCC 720
Qy 721 ATGGATCAGATTCAGCAGTGGCTTAAACAATCCGTAGAAATATTTGGTGAAAACCAGAAG 780
Db 721 ATGGATCAGATTCAGCAGTGGCTTAAACAATCCGTAGAAATATTTGGTGAAAACCAGAAG 780
Qy 781 GTAACGGCGGAACTGCAAAAAGCCATGTCTTCTGCGGTACAGCAAAATGCGGATGCTTCG 840
Db 781 GTAACGGCGGAACTGCAAAAAGCCATGTCTTCTGCGGTACAGCAAAATGCGGATGCTTCG 840
Qy 841 CGTTTTATTCTGCGCCAGAGTCGCGCATAAAAACTGCCAAAATAAAGGGAGAAAAATATG 900
Db 841 CGTTTTATTCTGCGCCAGAGTCGCGCATAAAAACTGCCAAAATAAAGGGAGAAAAATATG 900
Qy 901 TTAATTAGTAATGTGGGAATAAATCCCGCCGCTTATTTAAATAATCATTCTGTTGAGAAT 960
Db 901 TTAATTACTAATGTGGGAATAAATCCCGCCGCTTATTTAAATAATCATTCTGTTGAGAAT 960
Qy 961 AGTTCACAGACAGCTTCGCAATCCGTTAGCGCTAAAGATATTCTGAATAGTATTGGTATT 1020
Db 961 AGTTCACAGACAGCTTCGCAATCCGTTAGCGCTAAAGATATTCTGAATAGTATTGGTATT 1020
Qy 1021 AGCAGCAGTAAAGTCAGTGACCTGGGGTTGAGTCCTACACTGAGCGCGCCTGCGCCAGGG 1080
Db 1021 AGCAGCAGTAAAGTCAGTGACCTGGGGTTGAGTCCTACACTGAGCGCGCCTGCGCCAGGG 1080
Qy 1081 GTATTAACGCAAACCCCCCGGAACGATCACGTCCTTTTTAAAAGCCAGTATTCAAAATACC 1140
Db 1081 GTATTAACGCAAACCCCCGGAACGATCACGTCCTTTTTAAAAGCCAGTATTCAAAATACC 1140
Qy 1141 GACATGAATCAGGATTTGAATGCTCTGGCAAATAATGTCACGACTAAAGCGAATGAGGTT 1200
Db 1141 GACATGAATCAGGATTTGAATGCTCTGGCAAATAATGTCACGACTAAAGCGAATGAGGTT 1200
Qy 1201 GTGCAAACCCAGTTACGCGAGCAGCAGGCAGAAGTCGGAAAGTTTTTTGATATTAGCGGA 1260
Db 1201 GTGCAAACCCAGTTACGCGAGCAGCAGGCAGAAGTCGGAAAGTTTTTTGATATTAGCGGA 1260
Qy 1261 ATGTCTTCCAGTGCCGTTGCGCTGTTGGCTGCCGCGAATACGTTAATGCTGACGTTGAAC 1320
Db 1261 ATGTCTTCCAGTGCCGTTGCGCTGTTGGCTGCCGCGAATACGTTAATGCTGACGTTGAAC 1320
Qy 1321 CAGGCTGATAGCAAACTGTCTGGTAAGTTGTCATTAGTCAGTTTTGATGCAGCTAAAACG 1380
Db 1321 CAGGCTGATAGCAAACTGTCTGGTAAGTTGTCATTAGTCAGTTTTGATGCAGCTAAAACG 1380
Qy 1381 ACGGCAAGCTCCATGATGCGCGAAGGGATGAATGCGTTGTCCGGTAGTATTTCCCAGAGC 1440
Db 1381 ACGGCAAGCTCCATGATGCGCGAAGGGATGAATGCGTTGTCCGGTAGTATTTCCCAGAGC 1440
Qy 1441 GCGCTTCAGTTGGGGATCACTGGCGTGGGCGCCAAACTGGAATATAAGGGGCTGCAGAAT 1500
Db 1441 GCGCTTCAGTTGGGGATCACTGGCGTGGGCGCCAAACTGGAATATAAGGGGCTGCAGAAT 1500
Qy 1501 GAAAGAGGCGCGCTTAAACATAATG-CCGCGAAGATCGATAAACTGACCACTGAAAGCCA 1559
Db 1501 GAAAGAGGCGCGCTTAAACATAATGCCCGCGAAGATCGATAAACTGACCACTGAAAGCCA 1560
Qy 1560 CAGTATTAAAAACGTGCTGAACGGGCAGAATAGCGTCAAACTCGGTGCTGAAGGCGTCGA 1619
Db 1561 CAGTATTAAAAACGTGC-GAACGGGCAGAATAGCGTCAAACTCGGTGCTGAAGGCGTCGA 1619
Qy 1620 TTCTCTGAAATCGTTAAATATGAAGAAAACCGGTACCGATGCGACGAAAAATCTTAATGA 1679
Db 1620 TTCTCTGAAATCGTTAAATATGAAGAAAACCGGTACCGATGCGACGAAAAATCTTAATGA 1679
Qy 1680 TGCGACGCTTAAATCTAATGCCGGAACCAGCGCCACGGAAAGTCTGGGTATTAAAGACAG 1739
Db 1680 TGCGACGCTTAAATCTAATGCCGGAACCAGCGCCACGGAAAGTCTGGGTATTAAAGACAG 1739
Qy 1740 TAATAAACAAATCTCCCCTGAACATCAGGCTATTCTGTCGAAACGTCTTGAGTCTGTCGA 1799
Db 1740 TAATAAACAAATCTCCCCTGAACATCAGGCTATTCTGTCGAAACGTCTTGAGTCTGTCGA 1799
Qy 1800 ATCCGATATTCGTCTTGAGCAGAATACCATGGATATGACCCGAATCGATGCGCGCAAGAT 1859
Db 1800 ATCCGATATTCGTCTTGAGCAGAATACCATGGATATGACCCGAATCGATGCGCGCAAGAT 1859
Qy 1860 GCAGATGACGGGCGATCTGATTATGAAGAACTCGGTCACGGTCGGTGGTATTGCAGGGGC 1919
Db 1860 GCAGATGACGGGCGATCTGATTATGAAGAACTCGGTCACGGTCGGTGGTATTGCAGGGGC 1919
Qy 1920 GTCCGGGCAGTACGCCGCTACTCAGGAACGTTCCGAGCAGCAAATTAGCCAGGTGAATAA 1979
Db 1920 GTCCGGGCAGTACGCCGCTACTCAGGAACGTTCCGAGCAGCAAATTAGCCAGGTGAATAA 1979
Qy 1980 CCGGGTTGCCAGCACCGCATCGGACGAAGCCCGTGAAAGTTCACGTAAATCGACCAGCCT 2039
Db 1980 CCGGGTTGCCAGCACCGCATCGGACGAAGCCCGTGAAAGTTCACGTAAATCGACCAGCCT 2039
Qy 2040 GATTCAGGAAATGCTGAAAACAATGGAGAGCATTAACCAGTCGAAAGCATCCGCACTCGC 2099
Db 2040 GATTCAGGAAATGCTGAAAACAATGGAGAGCATTAACCAGTCGAAAGCATCCGCACTCGC 2099
Qy 2100 TGCTATCGCAGGCAATATTCGCGCTTAATCTGAAAGGTCATCTATACGCCATCATGGGTG 2159
Db 2100 TGCTATCGCAGGCAATATTCGCGCTTAATCTGAAAGGTCATCTATACGCCATCATGGGTG 2159
Qy 2160 TGATTTAATCGCGCTCCTGATGGCGAACTGGGGATATTATGCTTAATATTCAAAATTATT 2219
Db 2160 TGATTTAATCGCGCTCCTGATGGCGAACTGGGGATATTATGCTTAATATTCAAAATTATT 2219
Qy 2220 CCGCTTCTCCTCATCCGGGGATCGTTGCCGAACGGCCGCAGACTCCCTCGGCGAGCGAGC 2279
Db 2220 CCGCTTCTCCTCATCCGGGGATCGTTGCCGAACGGCCGCAGACTCCCTCGGCGAGCGAGC 2279
Qy 2280 ACGTCGAGACTGCCGTGGTACCGTCTACCACAGAACATCGCGGTACAGATATCATTTCAT 2339

```
Db   2280 ACGTCGAGACTGCCGTGGTACCGTCTACCACAGAACATCGCGGTACAGATATCATTTCAT 2339

Qy   2340 TATCGCAGGCGGCTACTAAAATCCACCAGGCACAGCAGACGCTGCAGTCAACGCCACCGA 2399
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2340 TATCGCAGGCGGCTACTAAAATCCACCAGGCACAGCAGACGCTGCAGTCAACGCCACCGA 2399

Qy   2400 TCTCTGAAGAGAATAATGACGAGCGCACGCTGGCGCGCCAGCAGTTGACCAGCAGCCTGA 2459
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2400 TCTCTGAAGAGAATAATGACGAGCGCACGCTGGCGCGCCAGCAGTTGACCAGCAGCCTGA 2459

Qy   2460 ATGCGCTGGCGAAGTCCGGCGTGTCATTATCCGCAGAACAAAATGAGAACCTGCGGAGCG 2519
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2460 ATGCGCTGGCGAAGTCCGGCGTGTCATTATCCGCAGAACAAAATGAGAACCTGCGGAGCG 2519

Qy   2520 CGTTTTCTGCGCCGACGTCGGCCTTATTTAGCGCTTCGCCTATGGCGCAGCCGAGAACAA 2579
          |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db   2520 CGTTTTCTGCGCCGACGTCGGCCTTA-TTAGCGCTTCGCCTATGGCGCAGCCGAGAACAA 2578

Qy   2580 CCATTTCTGATGCTGAGATTTGGGATATGGTTTCCCAAAATATATCGGCGATAGGTGACA 2639
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2579 CCATTTCTGATGCTGAGATTTGGGATATGGTTTCCCAAAATATATCGGCGATAGGTGACA 2638

Qy   2640 GCTATCTGGGCGTTTATGAAAACGTTGTCGCAGTCTATACCGATTTTTATCAGGCCTTCA 2699
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2639 GCTATCTGGGCGTTTATGAAAACGTTGTCGCAGTCTATACCGATTTTTATCAGGCCTTCA 2698

Qy   2700 GTGATATTCTTTCCAAAATGGGAGGCTGGTTATTACCAGGTAAGGACGGTAATACCGTTA 2759
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2699 GTGATATTCTTTCCAAAATGGGAGGCTGGTTATTACCAGGTAAGGACGGTAATACCGTTA 2758

Qy   2760 AGCTAGATGTTACCTCACTCAAAAATGATTTAAACAGTTTAGTCAATAAATATAATCAAA 2819
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2759 AGCTAGATGTTACCTCACTCAAAAATGATTTAAACAGTTTAGTCAATAAATATAATCAAA 2818

Qy   2820 TAAACAGTAATACCGTTTTATTTCCAGCGCAGTCAGGCAGCGGCGTTAAAGTAGCCACTG 2879
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db   2819 TAAACAGTAATAGCGTTTTATTTCCAGCGCAGTCAGGCAGCGGCGTTAAAGTAGCCACTG 2878

Qy   2880 AAGCGGAAGCGAGACAGTGGCTCAGTGAATTGAATTTACCGAATAGCTGCCTGAAATCTT 2939
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2879 AAGCGGAAGCGAGACAGTGGCTCAGTGAATTGAATTTACCGAATAGCTGCCTGAAATCTT 2938

Qy   2940 ATGGATCCGGTTATGTCGTCACCGTTGATCTGACGCCATTACAAAAAATGGTTCAGGATA 2999
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2939 ATGGATCCGGTTATGTCGTCACCGTTGATCTGACGCCATTACAAAAAATGGTTCAGGATA 2998

Qy   3000 TTGATGGTTTAGGCGCGCCGGGAAAAGACTCAAAACTCGAAATGGATAACGCCAAATATC 3059
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2999 TTGATGGTTTAGGCGCGCCGGGAAAAGACTCAAAACTCGAAATGGATAACGCCAAATATC 3058

Qy   3060 AAGCCTGGCAGTCGGGTTTTAAAGCGCAGGAAGAAAATATGAAAACCACATTACAGACGC 3119
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3059 AAGCCTGGCAGTCGGGTTTTAAAGCGCAGGAAGAAAATATGAAAACCACATTACAGACGC 3118

Qy   3120 TGACGCAAAAATATAGCAATGCCAATTCATTGTACGACAACCTGGTAAAAGTGCTGAGCA 3179
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3119 TGACGCAAAAATATAGCAATGCCAATTCATTGTACGACAACCTGGTAAAAGTGCTGAGCA 3178

Qy   3180 GTACGATAAGTAGCAGCCTGGAAACCGCCAAAAGCTTCCTGCAAGGATAACAGAAGAGGA 3239
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3179 GTACGATAAGTAGCAGCCTGGAAACCGCCAAAAGCTTCCTGCAAGGATAACAGAAGAGGA 3238

Qy   3240 TATTAATAATGGTTACAAGTGTAAGGACTCAG-CCCCCCGTCATAATGCCAGGTATGCAG 3298
          |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db   3239 TATTAATAATGGTTACAAGTGTAAGGACTCAGCCCCCCCCGTCATAATGCCAGGTATGCAG 3298

Qy   3299 ACCGAGATCAAAACGCAGGCCACGAATCTTGCGGCGAATCTTTCCGCAGTCAGAGAAAGT 3358
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3299 ACCGAGATCAAAACGCAGGCCACGAATCTTGCGGCGAATCTTTCCGCAGTCAGAGAAAGT 3358

Qy   3359 GCCACAGCGACGCTGTCAGGGGAAATTAAAGGCCCGCAACTGGAAGATTTTCCCGCGCTG 3418
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3359 GCCACAGCGACGCTGTCAGGGGAAATTAAAGGCCCGCAACTGGAAGATTTTCCCGCGCTG 3418
```

```
Qy   3419 ATCAAACAGGCGAGTCTGGATGC 3441
          |||||||||||||||||||||||
Db   3419 ATCAAACAGGCGAGTCTGGATGC 3441
```

RESULT 2
T67029
ID T67029 standard; DNA; 1230 BP.
AC T67029;
DT 04-AUG-1997 (first entry)
DE Salmonella secreted protein SspC DNA.
KW Salmonella secreted protein; SspC; bacterial-mediated endocytosis;
KW diagnosis; therapy; vaccine; attenuation; virulence; ss.
OS Salmonella typhimurium.
FH Key Location/Qualifiers
FT cds 1. .1230
FT /*tag= a
FT /product= SspC
FT /note= "the SspC amino acid sequence deduced
FT from the coding sequence does not fully
FT correspond to the SspC sequence given in
FT Fig 13"
PN WO9718225-A1.
PD 22-MAY-1997.
PF 14-NOV-1996; U18504.
PR 14-NOV-1995; US-006733.
PA (GEHO) GEN HOSPITAL CORP.
PI Miller SI;
DR WPI; 97-289217/26.
DR P-PSDB; W15269.
PT New isolated Salmonella secreted proteins and related genes - used
PT to develop products for the detection, treatment or prevention of
PT Salmonella infections
PS Claim 5; Fig 20; 95pp; English.
CC A DNA molecule (T67029) codes for Salmonella secreted protein SspC
CC (W15269), a virulence factor involved in bacterial mediated
CC endocytosis, which is able to translocate into human epithelial
CC cells. To identify Ssp genes, S. typhimurium mutants were
CC investigated in invasion assays on HEp-2 epithelial cells. 2
CC Ssp mutants with transposon insertions located between spaT and
CC prgH were identified. Of these, one mutant lacked Ssp of 87,
CC 42 and 36 kDa. The invasiveness of this mutant implicated the
CC 42 and 36 kDa Ssp in invasion. DNA sequencing of this region (see
CC also T67035) identified 2 complete and 2 partial open reading
CC frames, designated sspB, sspC (T67029), sspD (T67030) and sspA
CC (T67031). These nucleic acids can be used in methods for diagnosing
CC Salmonella infection and in methods of targetting an antigen to
CC epithelial cells and of inducing a cytotoxic T cell immune response
CC in a mammal. Ssp gene mutation can be used to create attenuated
CC Salmonella strains useful as live vaccines.
SQ Sequence 1230 BP; 362 A; 270 C; 317 G; 281 T;

Query Match 35.7%; Score 1230; DB 1; Length 1230;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1230; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    898 ATGTTAATTAGTAATGTGGGAATAAATCCCGCCGCTTATTTAAATAATCATTCTGTTGAG 957
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGTTAATTAGTAATGTGGGAATAAATCCCGCCGCTTATTTAAATAATCATTCTGTTGAG 60

Qy    958 AATAGTTCACAGACAGCTTCGCAATCCGTTAGCGCTAAAGATATTCTGAATAGTATTGGT 1017
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AATAGTTCACAGACAGCTTCGCAATCCGTTAGCGCTAAAGATATTCTGAATAGTATTGGT 120

Qy   1018 ATTAGCAGCAGTAAAGTCAGTGACCTGGGGTTGAGTCCTACACTGAGCGCGCCTGCGCCA 1077
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ATTAGCAGCAGTAAAGTCAGTGACCTGGGGTTGAGTCCTACACTGAGCGCGCCTGCGCCA 180

Qy   1078 GGGGTATTAACGCAAACCCCCGGAACGATCACGTCCTTTTTAAAAGCCAGTATTCAAAAT 1137
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GGGGTATTAACGCAAACCCCCGGAACGATCACGTCCTTTTTAAAAGCCAGTATTCAAAAT 240
```

```
Qy   1138 ACCGACATGAATCAGGATTTGAATGCTCTGGCAAATAATGTCACGACTAAAGCGAATGAG 1197
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ACCGACATGAATCAGGATTTGAATGCTCTGGCAAATAATGTCACGACTAAAGCGAATGAG 300

Qy   1198 GTTGTGCAAACCCAGTTACGCGAGCAGCAGGCAGAAGTCGGAAAGTTTTTTGATATTAGC 1257
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GTTGTGCAAACCCAGTTACGCGAGCAGCAGGCAGAAGTCGGAAAGTTTTTTGATATTAGC 360

Qy   1258 GGAATGTCTTCCAGTGCCGTTGCGCTGTTGGCTGCCGCGAATACGTTAATGCTGACGTTG 1317
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GGAATGTCTTCCAGTGCCGTTGCGCTGTTGGCTGCCGCGAATACGTTAATGCTGACGTTG 420

Qy   1318 AACCAGGCTGATAGCAAACTGTCTGGTAAGTTGTCATTAGTCAGTTTTGATGCAGCTAAA 1377
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AACCAGGCTGATAGCAAACTGTCTGGTAAGTTGTCATTAGTCAGTTTTGATGCAGCTAAA 480

Qy   1378 ACGACGGCAAGCTCCATGATGCGCGAAGGGATGAATGCGTTGTCCGGTAGTATTTCCCAG 1437
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 ACGACGGCAAGCTCCATGATGCGCGAAGGGATGAATGCGTTGTCCGGTAGTATTTCCCAG 540

Qy   1438 AGCGCGCTTCAGTTGGGGATCACTGGCGTGGGCGCCAAACTGGAATATAAGGGGCTGCAG 1497
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 AGCGCGCTTCAGTTGGGGATCACTGGCGTGGGCGCCAAACTGGAATATAAGGGGCTGCAG 600

Qy   1498 AATGAAAGAGGCGCGCTTAAACATAATGCCGCGAAGATCGATAAACTGACCACTGAAAGC 1557
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 AATGAAAGAGGCGCGCTTAAACATAATGCCGCGAAGATCGATAAACTGACCACTGAAAGC 660

Qy   1558 CACAGTATTAAAAACGTGCTGAACGGGCAGAATAGCGTCAAACTCGGTGCTGAAGGCGTC 1617
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 CACAGTATTAAAAACGTGCTGAACGGGCAGAATAGCGTCAAACTCGGTGCTGAAGGCGTC 720

Qy   1618 GATTCTCTGAAATCGTTAAATATGAAGAAAACCGGTACCGATGCGACGAAAAATCTTAAT 1677
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 GATTCTCTGAAATCGTTAAATATGAAGAAAACCGGTACCGATGCGACGAAAAATCTTAAT 780

Qy   1678 GATGCGACGCTTAAATCTAATGCCGGAACCAGCGCCACGGAAAGTCTGGGTATTAAAGAC 1737
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GATGCGACGCTTAAATCTAATGCCGGAACCAGCGCCACGGAAAGTCTGGGTATTAAAGAC 840

Qy   1738 AGTAATAAACAAATCTCCCCTGAACATCAGGCTATTCTGTCGAAACGTCTTGAGTCTGTC 1797
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 AGTAATAAACAAATCTCCCCTGAACATCAGGCTATTCTGTCGAAACGTCTTGAGTCTGTC 900

Qy   1798 GAATCCGATATTCGTCTTGAGCAGAATACCATGGATATGACCCGAATCGATGCGCGCAAG 1857
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 GAATCCGATATTCGTCTTGAGCAGAATACCATGGATATGACCCGAATCGATGCGCGCAAG 960

Qy   1858 ATGCAGATGACGGGCGATCTGATTATGAAGAACTCGGTCACGGTCGGTGGTATTGCAGGG 1917
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 ATGCAGATGACGGGCGATCTGATTATGAAGAACTCGGTCACGGTCGGTGGTATTGCAGGG 1020

Qy   1918 GCGTCCGGGCAGTACGCCGCTACTCAGGAACGTTCCGAGCAGCAAATTAGCCAGGTGAAT 1977
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 GCGTCCGGGCAGTACGCCGCTACTCAGGAACGTTCCGAGCAGCAAATTAGCCAGGTGAAT 1080

Qy   1978 AACCGGGTTGCCAGCACCGCATCGGACGAAGCCCGTGAAAGTTCACGTAAATCGACCAGC 2037
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 AACCGGGTTGCCAGCACCGCATCGGACGAAGCCCGTGAAAGTTCACGTAAATCGACCAGC 1140

Qy   2038 CTGATTCAGGAAATGCTGAAAACAATGGAGAGCATTAACCAGTCGAAAGCATCCGCACTC 2097
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 CTGATTCAGGAAATGCTGAAAACAATGGAGAGCATTAACCAGTCGAAAGCATCCGCACTC 1200

Qy   2098 GCTGCTATCGCAGGCAATATTCGCGCTTAA 2127
          |||||||||||||||||||||||||||||
Db   1201 GCTGCTATCGCAGGCAATATTCGCGCTTAA 1230


RESULT   3
T67030
ID   T67030 standard; DNA; 1032 BP.
```

```
AC   T67030;
DT   04-AUG-1997  (first entry)
DE   Salmonella secreted protein SspD DNA.
KW   Salmonella secreted protein; SspD; bacterial-mediated endocytosis;
KW   diagnosis; therapy; vaccine; attenuation; virulence; ss.
OS   Salmonella typhimurium.
FH   Key             Location/Qualifiers
FT   cds             1. .1032
FT                   /*tag= a
FT                   /product= SspD
FT                   /note= "the SspD amino acid sequence deduced
FT                   from the coding sequence does not fully
FT                   correspond to the SspD sequence given in
FT                   Fig 14"
PN   WO9718225-A1.
PD   22-MAY-1997.
PF   14-NOV-1996; U18504.
PR   14-NOV-1995; US-006733.
PA   (GEHO ) GEN HOSPITAL CORP.
PI   Miller SI;
DR   WPI; 97-289217/26.
DR   P-PSDB; W15270.
PT   New isolated Salmonella secreted proteins and related genes - used
PT   to develop products for the detection, treatment or prevention of
PT   Salmonella infections
PS   Claim 7; Fig 21; 95pp; English.
CC   A DNA molecule (T67030) codes for Salmonella secreted protein SspD
CC   (W15270), a virulence factor involved in bacterial mediated
CC   endocytosis, which is able to translocate into human epithelial
CC   cells.  To identify Ssp genes, S. typhimurium mutants were
CC   investigated in invasion assays on HEp-2 epithelial cells.  2
CC   Ssp mutants with transposon insertions located between spaT and
CC   prgH were identified.  Of these, one mutant lacked Ssp of 87,
CC   42 and 36 kDa.  The invasiveness of this mutant implicated the
CC   42 and 36 kDa Ssp in invasion.  DNA sequencing of this region (see
CC   also T67035) identified 2 complete and 2 partial open reading
CC   frames, designated sspB, sspC (T67029), sspD (T67030) and sspA
CC   (T67031).  These nucleic acids can be used in methods for diagnosing
CC   Salmonella infection and in methods of targetting an antigen to
CC   epithelial cells and of inducing a cytotoxic T cell immune response
CC   in a mammal.  Ssp gene mutation can be used to create attenuated
CC   Salmonella strains useful as live vaccines.
SQ   Sequence   1032 BP;    304 A;    242 C;    246 G;    240 T;
```

```
  Query Match             29.9%;  Score 1030.4;  DB 1;  Length 1032;
  Best Local Similarity   99.9%;  Pred. No. 7.8e-300;
  Matches 1031;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy   2198 ATGCTTAATATTCAAAATTATTCCGCTTCTCCTCATCCGGGGATCGTTGCCGAACGGCCG 2257
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGCTTAATATTCAAAATTATTCCGCTTCTCCTCATCCGGGGATCGTTGCCGAACGGCCG 60

Qy   2258 CAGACTCCCTCGGCGAGCGAGCACGTCGAGACTGCCGTGGTACCGTCTACCACAGAACAT 2317
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CAGACTCCCTCGGCGAGCGAGCACGTCGAGACTGCCGTGGTACCGTCTACCACAGAACAT 120

Qy   2318 CGCGGTACAGATATCATTTCATTATCGCAGGCGGCTACTAAAATCCACCAGGCACAGCAG 2377
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CGCGGTACAGATATCATTTCATTATCGCAGGCGGCTACTAAAATCCACCAGGCACAGCAG 180

Qy   2378 ACGCTGCAGTCAACGCCACCGATCTCTGAAGAGAATAATGACGAGCGCACGCTGGCGCGC 2437
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ACGCTGCAGTCAACGCCACCGATCTCTGAAGAGAATAATGACGAGCGCACGCTGGCGCGC 240

Qy   2438 CAGCAGTTGACCAGCAGCCTGAATGCGCTGGCGAAGTCCGGCGTGTCATTATCCGCAGAA 2497
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CAGCAGTTGACCAGCAGCCTGAATGCGCTGGCGAAGTCCGGCGTGTCATTATCCGCAGAA 300

Qy   2498 CAAAATGAGAACCTGCGGAGCGCGTTTTCTGCGCCGACGTCGGCCTTATTTAGCGCTTCG 2557
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CAAAATGAGAACCTGCGGAGCGCGTTTTCTGCGCCGACGTCGGCCTTATTTAGCGCTTCG 360
```

```
Qy    2558 CCTATGGCGCAGCCGAGAACAACCATTTCTGATGCTGAGATTTGGGATATGGTTTCCCAA 2617
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 CCTATGGCGCAGCCGAGAACAACCATTTCTGATGCTGAGATTTGGGATATGGTTTCCCAA 420

Qy    2618 AATATATCGGCGATAGGTGACAGCTATCTGGGCGTTTATGAAAACGTTGTCGCAGTCTAT 2677
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 AATATATCGGCGATAGGTGACAGCTATCTGGGCGTTTATGAAAACGTTGTCGCAGTCTAT 480

Qy    2678 ACCGATTTTTATCAGGCCTTCAGTGATATTCTTTCCAAAATGGGAGGCTGGTTATTACCA 2737
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 ACCGATTTTTATCAGGCCTTCAGTGATATTCTTTCCAAAATGGGAGGCTGGTTATTACCA 540

Qy    2738 GGTAAGGACGGTAATACCGTTAAGCTAGATGTTACCTCACTCAAAAATGATTTAAACAGT 2797
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 GGTAAGGACGGTAATACCGTTAAGCTAGATGTTACCTCACTCAAAAATGATTTAAACAGT 600

Qy    2798 TTAGTCAATAAATATAATCAAATAAACAGTAATACCGTTTTATTTCCAGCGCAGTCAGGC 2857
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 TTAGTCAATAAATATAATCAAATAAACAGTAATACCGTTTTATTTCCAGCGCAGTCAGGC 660

Qy    2858 AGCGGCGTTAAAGTAGCCACTGAAGCGGAAGCGAGACAGTGGCTCAGTGAATTGAATTTA 2917
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 AGCGGCGTTAAAGTAGCCACTGAAGCGGAAGCGAGACAGTGGCTCAGTGAATTGAATTTA 720

Qy    2918 CCGAATAGCTGCCTGAAATCTTATGGATCCGGTTATGTCGTCACCGTTGATCTGACGCCA 2977
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 CCGAATAGCTGCCTGAAATCTTATGGATCCGGTTATGTCGTCACCGTTGATCTGACGCCA 780

Qy    2978 TTACAAAAAATGGTTCAGGATATTGATGGTTTAGGCGCGCCGGGAAAAGACTCAAAACTC 3037
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 TTACAAAAAATGGTTCAGGATATTGATGGTTTAGGCGCGCCGGGAAAAGACTCAAAACTC 840

Qy    3038 GAAATGGATAACGCCAAATATCAAGCCTGGCAGTCGGGTTTTAAAGCGCAGGAAGAAAAT 3097
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 GAAATGGATAACGCCAAATATCAAGCCTGGCAGTCGGGTTTTAAAGCGCAGGAAGAAAAT 900

Qy    3098 ATGAAAACCACATTACAGACGCTGACGCAAAAATATAGCAATGCCAATTCATTGTACGAC 3157
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 ATGAAAACCACATTACAGACGCTGACGCAAAAATATAGCAATGCCAATTCATTGTACGAC 960

Qy    3158 AACCTGGTAAAAGTGCTGAGCAGTACGATAAGTAGCAGCCTGGAAACCGCCAAAAGCTTC 3217
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db     961 AACCTGGTAAAAGTGCTGAGCACTACGATAAGTAGCAGCCTGGAAACCGCCAAAAGCTTC 1020

Qy    3218 CTGCAAGGATAA 3229
           ||||||||||||
Db    1021 CTGCAAGGATAA 1032
```

```
RESULT   4
T67028
ID   T67028 standard; DNA; 870 BP.
AC   T67028;
DT   04-AUG-1997  (first entry)
DE   Salmonella secreted protein SspB DNA.
KW   Salmonella secreted protein; SspB; bacterial-mediated endocytosis;
KW   diagnosis; therapy; vaccine; attenuation; virulence; ss.
OS   Salmonella typhimurium.
FH   Key             Location/Qualifiers
FT   cds             1. .870
FT                   /*tag= a
FT                   /product= SspB (incomplete)
FT                   /note= "deduced amino acid residues from bases
FT                   286-354 of the coding sequence to not
FT                   correspond to the SspB sequence given in
FT                   Fig 22"
PN   WO9718225-A1.
PD   22-MAY-1997.
PF   14-NOV-1996; U18504.
PR   14-NOV-1995; US-006733.
PA   (GEHO ) GEN HOSPITAL CORP.
PI   Miller SI;
DR   WPI; 97-289217/26.
DR   P-PSDB; W15268.
PT   New isolated Salmonella secreted proteins and related genes - used
PT   to develop products for the detection, treatment or prevention of
PT   Salmonella infections
PS   Claim 3; Fig 19; 95pp; English.
CC   A DNA molecule (T67028) codes for (incomplete) Salmonella secreted
CC   protein SspB (W15268), a virulence factor involved in bacterial
CC   mediated endocytosis, which is able to translocate into human
CC   epithelial cells.  To identify Ssp genes, S. typhimurium mutants
CC   were investigated in invasion assays on HEp-2 epithelial cells.  2
CC   Ssp mutants with transposon insertions located between spaT and
CC   prgH were identified.  One mutant lacked Ssp of 87, 42 and 36 kDa.
CC   The invasiveness of this mutant implicated the 42 and 36 kDa Ssp in
CC   invasion.  DNA sequencing of this region (see also T67035)
CC   identified 2 complete and 2 partial open reading frames, designated
CC   sspB, sspC (T67029), sspD (T67030) and sspA (T67031).  These
CC   nucleic acids can be used in methods for diagnosing Salmonella
CC   infection and in methods of targetting an antigen to epithelial
CC   cells and of inducing a cytotoxic T cell immune response in a
CC   mammal.  Ssp gene mutation can be used to create attenuated
CC   Salmonella strains useful as live vaccines.
SQ   Sequence   870 BP;    216 A;    178 C;    281 G;    195 T;
```

```
  Query Match             25.2%;  Score 868.4;  DB 1;  Length 870;
  Best Local Similarity   99.9%;  Pred. No. 3.8e-251;
  Matches  869;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CGCAAAGCCGAGGAAACGAACCGCATTATGGGATGTATCGGGAAAGTCCTCGGCGCGCTG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CGCAAAGCCGAGGAAACGAACCGCATTATGGGATGTATCGGGAAAGTCCTCGGCGCGCTG 60

Qy      61 CTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCT 120

Qy     121 GCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCG 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCG 180

Qy     181 TTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTG 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTG 240

Qy     241 ATTGGCAAGGCGATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAG 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 ATTGGCAAGGCGATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAG 300

Qy     301 ATGGCCGGCAGCATTGTTGGTGCGATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTG 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 ATGGCCGGCAGCATTGTTGGTGCGATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTG 360

Qy     361 GTGGTCGCAGTTGTCGGGAAAGGCGCGGCGGCGAAACTGGGTAACGCGCTGAGCAAAATG 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GTGGTCGCAGTTGTCGGGAAAGGCGCGGCGGCGAAACTGGGTAACGCGCTGAGCAAAATG 420

Qy     421 ATGGGCGAAACGATTAAGAAGTTGGTGCCTAACGTGCTGAAACAGTTGGCGCAAAACGGC 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 ATGGGCGAAACGATTAAGAAGTTGGTGCCTAACGTGCTGAAACAGTTGGCGCAAAACGGC 480

Qy     481 AGCAAACTCTTTACCCAGGGGATGCAACGTATTACTAGCGGTCTGGGTAATGTGGGTAGC 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 AGCAAACTCTTTACCCAGGGGATGCAACGTATTACTAGCGGTCTGGGTAATGTGGGTAGC 540

Qy     541 AAGATGGGCCTGCAAACGAATGCCTTAAGTAAAGAGCTGGTAGGTAATACCCTAAATAAA 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 AAGATGGGCCTGCAAACGAATGCCTTAAGTAAAGAGCTGGTAGGTAATACCCTAAATAAA 600

Qy     601 GTGGCGTTGGGCATGGAAGTCACGAATACCGCAGCCCAGTCAGCCGGTGGTGTTGCCGAG 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 GTGGCGTTGGGCATGGAAGTCACGAATACCGCAGCCCAGTCAGCCGGTGGTGTTGCCGAG 660

Qy     661 GGCGTATTTATTAAAAATGCCAGCGAGGCGCTTGCTGATTTTATGCTCGCCCGTTTTGCC 720
```

```
Db    661 GGCGTATTTATTAAAAATGCCAGCGAGGCGCTTGCTGATTTTATGCTCGCCCGTTTTGCC 720

Qy    721 ATGGATCAGATTCAGCAGTGGCTTAAACAATCCGTAGAAATATTTGGTGAAAACCAGAAG 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 ATGGATCAGATTCAGCAGTGGCTTAAACAATCCGTAGAAATATTTGGTGAAAACCAGAAG 780

Qy    781 GTAACGGCGGAACTGCAAAAAGCCATGTCTTCTGCGGTACAGCAAAATGCGGATGCTTCG 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GTAACGGCGGAACTGCAAAAAGCCATGTCTTCTGCGGTACAGCAAAATGCGGATGCTTCG 840

Qy    841 CGTTTTATTCTGCGCCAGAGTCGCGCATAA 870
          |||||||||||||||||||||||||||||
Db    841 GGTTTTATTCTGCGCCAGAGTCGCGCATAA 870


RESULT   5
T67031
ID   T67031 standard; DNA; 194 BP.
AC   T67031;
DT   04-AUG-1997  (first entry)
DE   Salmonella secreted protein SspA DNA.
KW   Salmonella secreted protein; SspA; bacterial-mediated endocytosis;
KW   diagnosis; therapy; vaccine; attenuation; virulence; ss.
OS   Salmonella typhimurium.
FH   Key             Location/Qualifiers
FT   cds             1. .194
FT                   /*tag= a
FT                   /product= SspA (incomplete)
FT                   /note= "the translated sequence does not fully
FT                   correspond to the SspA amino acid sequence
FT                   given in Fig 23"
PN   WO9718225-A1.
PD   22-MAY-1997.
PF   14-NOV-1996; U18504.
PR   14-NOV-1995; US-006733.
PA   (GEHO ) GEN HOSPITAL CORP.
PI   Miller SI;
DR   WPI; 97-289217/26.
DR   P-PSDB; W15271.
PT   New isolated Salmonella secreted proteins and related genes - used
PT   to develop products for the detection, treatment or prevention of
PT   Salmonella infections
PS   Claim 9; Fig 22; 95pp; English.
CC   A DNA molecule (T67031) codes for (incomplete) Salmonella secreted
CC   protein SspA (W15271), a virulence factor able to translocate into
CC   human epithelial cells.  To identify Ssp genes, S. typhimurium
CC   mutants were examined in invasion assays on HEp-2 epithelial cells.
CC   2 Ssp mutants with transposon insertions located between spaT and
CC   prgH were identified.  Of these, one mutant lacked Ssp of 87,
CC   42 and 36 kDa.  The invasiveness of this mutant implicated the
CC   42 and 36 kDa Ssp in invasion.  DNA sequencing of this region (see
CC   also T67035) identified 2 complete and 2 partial open reading
CC   frames, designated sspB, sspC (T67029), sspD (T67030) and sspA
CC   (T67031).  These nucleic acids can be used in methods for diagnosing
CC   Salmonella infection and in methods of targetting an antigen to
CC   epithelial cells and of inducing a cytotoxic T cell immune response
CC   in a mammal.  Ssp gene mutation can be used to create attenuated
CC   Salmonella strains useful as live vaccines.
SQ   Sequence   194 BP;    55 A;    51 C;    53 G;    35 T;


  Query Match              5.6%;  Score 194;  DB 1;  Length 194;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-48;
  Matches  194;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   3248 ATGGTTACAAGTGTAAGGACTCAGCCCCCCGTCATAATGCCAGGTATGCAGACCGAGATC 3307
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGGTTACAAGTGTAAGGACTCAGCCCCCCGTCATAATGCCAGGTATGCAGACCGAGATC 60

Qy   3308 AAAACGCAGGCCACGAATCTTGCGGCGAATCTTTCCGCAGTCAGAGAAAGTGCCACAGCG 3367
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AAAACGCAGGCCACGAATCTTGCGGCGAATCTTTCCGCAGTCAGAGAAAGTGCCACAGCG 120

Qy   3368 ACGCTGTCAGGGGAAATTAAAGGCCCGCAACTGGAAGATTTTCCCGCGCTGATCAAACAG 3427
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ACGCTGTCAGGGGAAATTAAAGGCCCGCAACTGGAAGATTTTCCCGCGCTGATCAAACAG 180

Qy   3428 GCGAGTCTGGATGC 3441
          ||||||||||||||
Db    181 GCGAGTCTGGATGC 194


RESULT   6
T85939
ID   T85939 standard; DNA; 1743 BP.
AC   T85939;
DT   10-MAR-1998  (first entry)
DE   Shigella flexneri IpaB gene.
KW   IpaB; apoptosis; interleukin-1-beta converting enzyme; ICE; cacner;
KW   autoimmune disease; inflammation; AIDS; degenerative disease;
KW   Alzheimer's disease; myelodysplastic disorder; ischaemic injury;
KW   toxin-induced liver disease; gene therapy; shigellosis; ss.
OS   Shigella flexneri.
PN   WO9726790-A1.
PD   31-JUL-1997.
PF   10-JAN-1997; U00399.
PR   25-JAN-1996; US-591079.
PA   (UYNY ) UNIV NEW YORK STATE.
PI   Chen Y,  Zychlinsky A;
DR   WPI; 97-393276/36.
DR   P-PSDB; W29893.
PT   Shigella IpaB protein induces apoptosis by binding to
PT   interleukin-1-beta-converting enzyme - useful for treating cancer,
PT   auto-immunity, inflammation, etc. by eradication of unwanted cells
PS   Claim 3; Page 67-69; 108pp; English.
CC   This sequence comprises DNA encoding the IpaB protein (see W29893)
CC   of Shigella flexneri.  This protein induces apoptosis by binding to
CC   interleukin-1-beta converting enzyme (ICE).  A claimed method of
CC   inducing apoptosis in a eukaryotic cell, thereby killing the cell,
CC   comprises: (a) providing to a cell to be killed a DNA molecule (I)
CC   in expressible form which encodes the Shigella IpaB protein or a
CC   functional derivative; and (b) causing (I) to be expressed in the
CC   cell, thereby producing the IpaB protein or derivative which
CC   induces apoptosis and kills the cell.  This approach is useful in
CC   treating e.g. cancer, autoimmunity, inflammation and chronic viral
CC   infections.  Protein or peptide fragments, and the DNA encoding
CC   them, can also be used to treat diseases that involve apoptotic
CC   mechanisms in their pathogenesis, e.g. AIDS, degenerative diseases
CC   such as Alzheimer's disease, myelodysplastic disorders, ischaemic
CC   injury or toxin-induced liver disease.  Methods of gene therapy
CC   relying upon controlled expression of IpaB in a target cell are
CC   disclosed.  Also claimed are methods for: inhibiting the
CC   interaction of an apoptosis-inducing protein or peptide with ICE;
CC   detecting a compound capable of inhibiting the binding of IpaB to
CC   ICE; screening a candidate protein for its ability to interact
CC   with IpaB in a cell; and treatment of shigellosis.  Homologues of
CC   IpaB such as yopB of Yersinia entercolitica and Yersinia
CC   pseudotuberculosis, and sipB of Salmonella typhimurium and
CC   Salmonella typhi can also be used in the methods of the invention.
SQ   Sequence   1743 BP;    632 A;    360 C;    302 G;    449 T;


  Query Match              5.2%;  Score 178.6;  DB 1;  Length 1743;
  Best Local Similarity   58.3%;  Pred. No. 1.9e-43;
  Matches  313;  Conservative    0;  Mismatches  224;  Indels    0;  Gaps    0;

Qy      1 CGCAAAGCCGAGGAAACGAACCGCATTATGGGATGTATCGGGAAAGTCCTCGGCGCGCTG 60
          || ||||| || |||   ||| | | ||||| ||| | |||||| | || || || ||
Db    892 CGTAAAGCAGAAGAACTCAACAGAGTAATGGGTTGTGTTGGGAAAATACTTGGGGCACTT 951

Qy     61 CTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCT 120
          |||| || || || ||||| || || | |||| | || || ||   ||| || |||||
Db    952 TTAACTATCGTTAGTGTTGTTGCAGCAGCTTTTTCTGGAGGAGCCTCTCTAGCACTGGCA 1011

Qy    121 GCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCG 180
```

```
Db   1012 GCTGTTGGTTTAGCTCTTATGGTTACGGATGCTATAGTACAAGCAGCGACCGGCAATTCC 1071

Qy    181 TTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTG 240

Db   1072 TTCATGGAACAAGCCCTGAATCCGATCATGAAAGCAGTCATTGAACCCTTAATCAAACTC 1131

Qy    241 ATTGGCAAGGCGATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAG 300

Db   1132 CTTTCAGATGCATTTACAAAAATGCTCGAAGGCTTGGGCGTCGACTCGAAAAAAGCCAAA 1191

Qy    301 ATGGCCGGCAGCATTGTTGGTGCGATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTG 360

Db   1192 ATGATTGGCTCTATTCTGGGGGCAATCGCAGGCGCTCTTGTCCTAGTTGCAGCAGTCGTT 1251

Qy    361 GTGGTCGCAGTTGTCGGGAAAGGCGCGGCGGCGAAACTGGGTAACGCGCTGAGCAAAATG 420

Db   1252 CTCGTAGCCACTGTTGGTAAACAGGCAGCAGCAAAACTTGCAGAAAATATTGGCAAAATA 1311

Qy    421 ATGGGCGAAACGATTAAGAAGTTGGTGCCTAACGTGCTGAAACAGTTGGCGCAAAACGGC 480

Db   1312 ATAGGTAAAACCCTCACAGACCTTATACCAAAGTTTCTCAAGAATTTTTCTTCTCAACTG 1371

Qy    481 AGCAAACTCTTTACCCAGGGGATGCAACGTATTACTAGCGGTCTGGGTAATGTGGGT 537

Db   1372 GACGATTTAATCACTAATGCTGTTGCCAGATTAAATAAATTTCTTGGTGCAGCGGGT 1428
```

```
RESULT   7
Q25532/c
ID   Q25532 standard; DNA; 1648 BP.
AC   Q25532;
DT   22-NOV-1992  (first entry)
DE   Sequence of genomic clone contg. the entire Histidine-rich
DE   protein (HisRP) gene.
KW   Malaria vaccine; Histidine-rich protein; cytoadherence; ss.
OS   Plasmodium lophurae.
FH   Key             Location/Qualifiers
FT   signal_peptide  219. .359
FT                   /*tag= a
FT   cds             491. .562
FT                   /*tag= b
FT                   /product= propeptide
FT   intron          360. .490
FT                   /*tag= c
FT   misc_feature    60. .64
FT                   /*tag= d
FT                   /label= splice sequence
FT   misc_feature    485. .490
FT                   /*tag= e
FT                   /label= splice sequence
FT   mat_peptide     563. .1477
FT                   /*tag= f
PN   US5116965-A.
PD   25-MAY-1992.
PF   26-AUG-1986; 900401.
PR   26-AUG-1986; US-900401.
PA   (SLOK ) SLOAN KETTERING INST CANCER.
PI   Pologe L, Ravetch JV;
DR   WPI; 92-199590/24.
DR   P-PSDB; R24393.
PT   Histidine-rich protein associated with Plasmodium knob phenotype -
PT   and DNA encoding it, used for in vitro diagnosis of P.
PT   Falciparum infection.
PS   Disclosure; Fig 7A-B; 29pp; English.
CC   Two variants of HisRP are produced by P. falciparum.  One is
CC   associated with what is referred to as "knobby phenotype" (K30) and
CC   "knobless phenotype" (K-).  The "knobby" and "knobless" phenotypes
CC   have been implicated in cytoadherence, which is characteristic of
CC   erythrocyte infection.  It has now been found that cDNA expressing
CC   both K+ and K- HisRP can be obtained by the use of P. lophurae HisRP
CC   expressing DNA.  The genomic clone (Q25532) is encoded in two
CC   exons, separating the signal peptide-encoding sequence from the
CC   pro-sequence, confirming that synthesis of the protein occurs via
CC   the preproprotein.  Oligo. probes synthesised to the signal
CC   peptide-encoding exon reveal multiple homologous DNA sequences in
CC   the P. lophurae genome.  The sequence of mature proteins is arranged
CC   in numerous tandem repeats with up to nine histidine residues in a
CC   row, similar to other Plasmodium proteins for which sequence data
CC   have so far been reported.
SQ   Sequence   1648 BP;    569 A;    488 C;    150 G;    441 T;
```

```
  Query Match             1.4%;  Score 49.4;  DB 1;  Length 1648;
  Best Local Similarity   44.9%;  Pred. No. 0.00014;
  Matches  188;  Conservative    0;  Mismatches  231;  Indels    0;  Gaps    0;

Qy      29 TGGGATGTATCGGGAAAGTCCTCGGCGCGCTGCTAACCATTGTCAGCGTTGTGGCCGCTG 88

Db    1425 TGGTGTGCATCGTGATGATGATGGTGGTGGTGGTGTGCATCGTGATGATGATGGTGGTGG 1366

Qy      89 TTTTTACCGGTGGGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTGGCCG 148

Db    1365 TGATGTGCATCGTGGTGATGGTGGTGGTGGTGTGCATCGTGATGATGATGGTGGTGGTGG 1306

Qy     149 ATGAAATTGTGAAGGCGGCGACGGGAGTGTCGTTTATTCAGCAGGCGCTAAACCCGATTA 208

Db    1305 TGTGCATCGTGATGATGATGGTGGTGGTGGTGTGCATCGTGATGATGATGGTGGTGGTGA 1246

Qy     209 TGGAGCATGTGCTGAAGCCGTTAATGGAGCTGATTGGCAAGGCGATTACCAAAGCGCTGG 268

Db    1245 TGTGCATCGTGATCATGATGGTGGTGGTGGTGTGCATCGTGATGATGATGGTGGTGGTGG 1186

Qy     269 AAGGATTAGGCGTCGATAAGAAAACGGCAGAGATGGCCGGCAGCATTGTTGGTGCGATTG 328

Db    1185 TGTGCATCATGGTGGTGGTGGTGATGATGGTGGTGACCATGGTGGTGGTGGTGATGATGG 1126

Qy     329 TCGCCGCTATTGCCATGGTGGCGGTCATTGTGGTGGTCGCAGTTGTCGGGAAAGGCGCGG 388

Db    1125 TGGTGACCATGGTGGTGGTGGTGATGATGGTGGTGACCATGGTGGTGGTGGTGATGATGG 1066

Qy     389 CGGCGAAACTGGGTAACGCGCTGAGCAAAATGATGGGCGAAACGATTAAGAAGTTGGTG 447

Db    1065 TGTGGAGCATGGTGGTGGTGGTGATGATGGTGTGGAGCATGGTGGTGGTGGTGATGATG 1007
```

```
RESULT   8
Q55750
ID   Q55750 standard; DNA; 1505 BP.
AC   Q55750;
DT   23-JUL-1994  (first entry)
DE   Genomic clone G11F, includes a Brassica root specific promoter.
KW   Plants; transformation; food crops; toxins; insects; fungal;
KW   infection; ss.
OS   Brassica napus.
FH   Key             Location/Qualifiers
FT   misc_feature    552. .554
FT                   /*tag= a
FT                   /note= "translation start site"
FT   caat_signal     438. .440
FT                   /*tag= b
FT   tata_signal     485. .488
FT                   /*tag= c
PN   WO9402619-A.
PD   03-FEB-1994.
PF   12-JUL-1993; U06541.
PR   16-JUL-1992; US-915246.
PA   (PION-) PIONEER HI-BRED INT INC.
PI   Baszczynski CJ,  Bellemare G,  Boivin R,  Fallis L;
DR   WPI; 94-048877/06.
PT   Brassica root-specific promoter sequence and vectors containing
PT   it - for expression of e.g. toxins in the roots of crops
PT   susceptible to root diseases
PS   Disclosure; Fig 4; 54pp; English.
CC   The sequence shows the G11F genomic clone isolated from DNA from
CC   the roots of Brassica napus and containing a promoter sequence
```

```
CC   upstream from the translation start site.  The promoter enhances
CC   gene expression in roots and can be used to confer disease
CC   resistance or immunity upon a susceptible plant root by transforming
CC   the plant with a composite gene capable of abundant expression in the
CC   gene, comprising the promoter and the gene of interest.  This method
CC   can be used to transform plants, partic food crops that have roots
CC   susceptible to fungal or insect diseases.  The roots of the transformed
CC   plants express a toxin or substance that provides immunity or resistance
CC   to the fungal or insect diseases.
CC   See also Q55749 and Q56913-5.
SQ   Sequence   1505 BP;    395 A;    201 C;    528 G;    381 T;


  Query Match             1.3%;  Score 45.2;  DB 1;  Length 1505;
  Best Local Similarity   47.2%;  Pred. No. 0.0024;
  Matches  137;  Conservative    0;  Mismatches  153;  Indels    0;  Gaps    0;

Qy    111 GGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGAC 170
Db   1047 GGTGGTGGAGGCGGTGGAAATGGAGGCGGTGGAGGAGGTGGAGGTGCACACGGTGGTGGA 1106

Qy    171 GGGAGTGTCGTTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTT 230
Db   1107 TACGGTGGTGGAGAAGGCGCTGGTGCTGGAGGAGGATATGGAGGTGGCGGTGCAGGTGGA 1166

Qy    231 AATGGAGCTGATTGGCAAGGCGATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAAGAA 290
Db   1167 CATGGAGGTGGTGGAGGCGGGGGAAAAGGAGGCGGTGGAGGAGGAGGTTCTGGCGCCGGT 1226

Qy    291 AACGGCAGAGATGGCCGGCAGCATTGTTGGTGCGATTGTCGCCGCTATTGCCATGGTGGC 350
Db   1227 GGAGCTCACGGTGGTGGTTATGGTGCCGGAGGTGGAGCTGGAGAGGGATACGGTGGTGGT 1286

Qy    351 GGTCATTGTGGTGGTCGCAGTTGTCGGGAAAGGCGCGGCGGCGAAACTGG 400
Db   1287 GGTGGAGAAGGAGGACACGGTGGTGGAGGAGGCGGTGGTGGTGGAGCTGG 1336


RESULT   9
Q63879
ID   Q63879 standard; DNA; 4018 BP.
AC   Q63879;
DT   07-DEC-1994  (first entry)
DE   Polyhydroxyalkanoate decomposing activity polypeptide.
KW   Polyhydroxyalkanoate decomposing enzyme; PHA; ds.
OS   Zoogloea ramigera.
FH   Key             Location/Qualifiers
FT   cds             395. .1153
FT                   /*tag=  a
FT                   /product=
FT                   /note= "ORF 1"
FT   cds             3173. .3958
FT                   /*tag=  b
FT                   /product=
FT                   /note= "ORF 2"
PN   J06086681-A.
PD   29-MAR-1994.
PF   07-SEP-1992; 279099.
PR   07-SEP-1992; JP-279099.
PA   (ELED ) DENKI KAGAKU KOGYO KK.
PA   (UYKA-) GH KANAGAWA DAIGAKU.
PA   (CHIK-) ZH CHIKYU KANKYO SANGYO GIJITSU KENKYU.
DR   WPI; 94-140539/17.
DR   P-PSDB; R53915 AND R63805.
PT   DNA encoding poly-hydroxy-alkanoate decomposing enzyme - useful
PT   for preparing poly-hydroxy-alkanoate in large quantities
PS   Claim 3; Page 9-14; 15pp; Japanese.
CC   Q63879 shows two ORFs, the products of which show
CC   polyhydroxyalkanoate decomposing activity. The method used allows
CC   the polypeptide to be prepared in large quantities.
SQ   Sequence   4018 BP;    767 A;    1298 C;    1305 G;    648 T;


  Query Match             1.3%;  Score 43.6;  DB 1;  Length 4018;
  Best Local Similarity   44.6%;  Pred. No. 0.013;
  Matches  172;  Conservative    0;  Mismatches  214;  Indels    0;  Gaps    0;

Qy     64 ACCATTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCTGCG 123
Db    392 ACGATGACCCTCCCGGTCGCGGTGACTACGATGGCTGGCATGGGCACGGTCATGGCCACG 451

Qy    124 GTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCGTTT 183
Db    452 GTCACGGGCACGGCCATGGCGGCCACGGCCACGGTGGTGGCGGTCACGGCCACGGCGGCG 511

Qy    184 ATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTGATT 243
Db    512 GCGGTCATGACGGTGGAGGCCATGACGGTGGCGGTAACCACGGCGGTGGTGGTAACCATG 571

Qy    244 GGCAAGGCGATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAGATG 303
Db    572 GCGGCGGCGGCAACCACGGTGGTGGCGGCAACCATGGCGGTCGCGGCAACCATGGCGGTG 631

Qy    304 GCCGGCAGCATTGTTGGTGCGATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTGGTG 363
Db    632 GCGGCAACCACGGTGGTCGCGGCAACCACGGCGGTGGTGACAACCATGGCGGTGGCGGCA 691

Qy    364 GTCGCAGTTGTCGGGAAAGGCGCGGCGGCGAAACTGGGTAACGCGCTGAGCAAAATGATG 423
Db    692 ACCACCGTGGTGGCGGCAACCATGGCGGTGGCGGTAACCACGGTGGTGGCGGCAACCATG 751

Qy    424 GGCGAAACGATTAAGAAGTTGGTGCCC 449
Db    752 GCGGTGGCGGCAACCACGGTGGTGGC 777


RESULT  10
Q67401/c
ID   Q67401 standard; DNA; 10366 BP.
AC   Q67401;
DT   13-APR-1995  (first entry)
DE   Huntingtin DNA\IT15 gene.
KW   Polymerase chain reaction; primer; PCR; amplify; Huntingtin; IT15;
KW   open reading frame; polymorphic; (CAG)n; trinucleotide repeat;
KW   allele; Huntingdon's disease; HD; chromosome; juvenile HD; unstable;
KW   expandable; linkage disequilibrium; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             316. .9750
FT                   /*tag=  a
PN   EP-614977-A.
PD   14-SEP-1994.
PF   07-MAR-1994; 301587.
PR   05-MAR-1993; US-027498.
PR   01-JUL-1993; US-085000.
PA   (GEHO ) GEN HOSPITAL CORP.
PI   Ambrose CM,  Duyao MP,  Gusella JF,  MacDonald ME;
DR   WPI; 94-281205/35.
DR   P-PSDB; R58777.
PT   New huntingtin protein and related nucleic acid, antibodies etc.
PT   - for treatment and diagnosis of neuro-degenerative disease,
PT   specifically Huntington's disease
PS   Claim 5; Page 22-33; 66pp; English.
CC   This sequence represents a large gene, termed "Huntingtin" or
CC   "IT15" which is present in the proximal part of the 0.5 mb segment
CC   between members D4S180 and D4S182 on chromosome 4.  This gene
CC   spans about 210 kb and encodes a protein of approx. 348 kD.  The
CC   huntingtin open reading frame contains a polymorphic (CAG)n
CC   trinucleotide repeat with at least 17 alleles in the normal
CC   population, varying from about 11 to about 34 CAG copies.  On
CC   Huntingdon's disease (HD) chromosome, the length of the CAG repeat
CC   is substantially increased, with between 37 to at least 73 copies.
CC   This shows an apparent correlation with age of onset, the longest
CC   segments are detected in juvenile HD cases.  The presence of an
CC   unstable, expandable trinucleotide repeat on HD chromosomes in the
CC   region of the strongest linkage disequilibrium with the disorder
```

```
CC   suggest that this alteration underlies the dominant phenotype of HD.
SQ   Sequence   10366 BP;    2426 A;    2808 C;    2742 G;    2390 T;

  Query Match             1.1%;  Score 38.2;  DB 1;  Length 10366;
  Best Local Similarity   51.6%;  Pred. No. 1;
  Matches  111;  Conservative    0;  Mismatches  103;  Indels    1;  Gaps    1;

Qy      74 GCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTG 133
           ||   ||||| || |       | |  ||||||  |  |||   ||| || | ||   |
Db     552 GCCGGGTGGCGGCGGGGGCGGCGGCGGGGGCGGCTGCGGCTGAGGCAGCAGCGGCTGTGC 493

Qy     134 CGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCGTTTAT-TCAGCAG 192
           | |    || ||| || |||  || | ||||||||| |||  | | ||    |  | || |
Db     492 CTGCGGCGGCGGCTGAGGAAGCTGAGGAGGCGGCGGCGGCGGCGGCGGTGGCGGCTG 433

Qy     193 GCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTGATTGGCAAGGCG 252
             |||    | | |  ||  ||   |||||  || | |  ||  |||| |      | |
Db     432 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 373

Qy     253 ATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAA 287
           |     |||  || || ||| | |  | |  | |
Db     372 CTGCTGGAAGGACTTGAGGGACTCGAAGGCCTTCA 338


RESULT  11
T97924/c
ID   T97924 standard; DNA; 10366 BP.
AC   T97924;
DT   13-MAR-1998  (first entry)
DE   Composite sequence of a novel huntingtin (IT15) gene.
KW   Huntingin gene; IT15 gene; Huntington's disease; trinucleotide repeat;
KW   neurodegenerative disorder; HD; gene therapy; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             316. .9750
FT                   /*tag=  a
PN   US5686288-A.
PD   11-NOV-1997.
PF   20-MAY-1994; 246982.
PR   05-MAR-1993; US-027498.
PR   01-JUL-1993; US-085000.
PA   (GEHO ) GEN HOSPITAL CORP.
PI   Ambrose CM,  Duyao MP,  Gusella JF,  MacDonald ME;
DR   WPI; 97-558144/51.
DR   P-PSDB; W36887.
PT   Nucleic acid encoding huntingtin protein - useful for gene therapy
PT   of Huntington's disease
PS   Claim 3; Fig 4; 112pp; English.
CC   The present sequence represents a novel gene, termed huntingin or IT15.
CC   This gene encodes a previously undescribed protein. The huntingtin
CC   reading frame contains a polymorphic (CAG)n trinucleotide repeat with at
CC   least 17 alleles in the normal population, varying from about 11 to 34
CC   CAG copies. Huntington's disease (HD) is a progressive neurodegenerative
CC   disorder characterised by motor disturbance, cognitive loss and
CC   psychiatric manifestations. The genetic defect causing HD is assigned to
CC   chromosome 4. On HD chromosomes, the length of the trinucleotide CAG
CC   repeat is substantially increased, e.g. about 37 to at least 73 copies.
CC   The present nucleic acid and proteins encoded by it, may be used for the
CC   diagnosis or treatment of Huntington's disease. The huntingtin gene
CC   is especially used in gene therapy of a symptomatic or presymptomatic
CC   patient. The method comprises providing a functional huntingtin gene with
CC   a (CAG)n repeat of the normal range of 11-34 copies, or an antisense
CC   sequence, to the desired cells of the patient, in a manner that permits
CC   the expression of the huntingtin protein provided by the gene, or
CC   inhibits expression of the mutated huntingtin gene, for a time and in a
CC   quantity sufficient to provide the huntingtin function to the cells of
CC   the patient.
SQ   Sequence   10366 BP;    2426 A;    2809 C;    2742 G;    2389 T;

  Query Match             1.1%;  Score 38.2;  DB 1;  Length 10366;
  Best Local Similarity   51.6%;  Pred. No. 1;
  Matches  111;  Conservative    0;  Mismatches  103;  Indels    1;  Gaps    1;

Qy      74 GCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTG 133
           ||   ||||| || |       | |  ||||||  |  |||   ||| || | ||   |
Db     552 GCCGGGTGGCGGCGGGGGCGGCGGCGGGGGCGGCTGCGGCTGAGGCAGCAGCGGCTGTGC 493

Qy     134 CGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCGTTTAT-TCAGCAG 192
           | |    || ||| || |||  || | ||||||||| |||  | | ||    |  | || |
Db     492 CTGCGGCGGCGGCTGAGGAAGCTGAGGAGGCGGCGGCGGCGGCGGCGGTGGCGGCTG 433

Qy     193 GCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTGATTGGCAAGGCG 252
             |||    | | |  ||  ||   |||||  || | |  ||  |||| |      | |
Db     432 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 373

Qy     253 ATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAA 287
           |     |||  || || ||| | |  | |  | |
Db     372 CTGCTGGAAGGACTTGAGGGACTCGAAGGCCTTCA 338


RESULT  12
V05828/c
ID   V05828 standard; DNA; 10366 BP.
AC   V05828;
DT   01-JUN-1998  (first entry)
DE   Human huntingtin coding sequence.
KW   Human; huntingtin gene; Huntington's disease; chromosome; marker;
KW   locus; antisense; gene therapy; diagnosis; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             316. .9750
FT                   /*tag=  a
FT                   /product= huntingtin_protein
PN   US5693757-A.
PD   02-DEC-1997.
PF   30-MAY-1995; 453265.
PR   20-MAY-1994; US-246982.
PR   05-MAR-1993; US-027498.
PR   01-JUL-1993; US-085000.
PR   30-MAY-1995; US-453265.
PA   (GEHO ) GEN HOSPITAL CORP.
PI   Ambrose CM, Duyao MP, Gusella JF, MacDonald ME;
DR   WPI; 98-031815/03.
DR   P-PSDB; W44742.
PT   Huntingtin protein and related nucleic acid - for diagnosis or
PT   therapy of Huntington's disease
PS   Example 1; Fig 4; 112pp; English.
CC   This is the nucleotide sequence which encodes the human huntingtin
CC   protein.  The gene sequence is characterised in that it contains a
CC   number of CAG repeats in the 5' region (in this case 23 repeats).
CC   In healthy individuals the usual number of CAG repeats in the gene
CC   is 11-34, whereas in patients suffering from Huntington's disease (HD),
CC   the number of repeats increases to 37-73 or 37-86.  The huntingtin gene
CC   spans 210 kb and encodes a protein of 348 kD. The gene is found in a
CC   500 kb region between the chromosomal markers D4S180 and D4S182 and is
CC   preferentially mapped to the locus 4p16.3.  The protein or the gene
CC   encoding it, is useful for detecting a predisposition to develop HD,
CC   for diagnosis and treatment of HD, especially by antisense and gene
CC   therapy.
SQ   Sequence   10366 BP;    2426 A;    2809 C;    2742 G;    2389 T;

  Query Match             1.1%;  Score 38.2;  DB 1;  Length 10366;
  Best Local Similarity   51.6%;  Pred. No. 1;
  Matches  111;  Conservative    0;  Mismatches  103;  Indels    1;  Gaps    1;

Qy      74 GCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTG 133
           ||   ||||| || |       | |  ||||||  |  |||   ||| || | ||   |
Db     552 GCCGGGTGGCGGCGGGGGCGGCGGCGGGGGCGGCTGCGGCTGAGGCAGCAGCGGCTGTGC 493

Qy     134 CGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCGTTTAT-TCAGCAG 192
           | |    || ||| || |||  || | ||||||||| |||  | | ||    |  | || |
Db     492 CTGCGGCGGCGGCTGAGGAAGCTGAGGAGGCGGCGGCGGCGGCGGCGGTGGCGGCTG 433
```

```
Qy    193 GCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTGATTGGCAAGGCG 252
          ||| |  | | |  ||  ||  |||||  || | |  ||  |||| |      |  |
Db    432 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 373

Qy    253 ATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAA 287
          |    |||  || || ||| | |  | |  | |
Db    372 CTGCTGGAAGGACTTGAGGGACTCGAAGGCCTTCA 338


RESULT  13
X18930/c
ID   X18930 standard; DNA; 515 BP.
AC   X18930;
DT   11-MAY-1999  (first entry)
DE   Human huntingtin gene.
KW   Huntingtin protein; osm-10 gene; regulatory region; TREND disease;
KW   Huntington's disease; triplet repeat neurological disease;
KW   transgenic nematode; ss.
OS   Homo sapiens.
PN   WO9902652-A1.
PD   21-JAN-1999.
PF   10-JUL-1998; U14376.
PR   11-JUL-1997; US-052324.
PA   (GEHO ) GEN HOSPITAL CORP.
PI   Faber P, Hart AC, MacDonald M;
DR   WPI; 99-120856/10.
DR   P-PSDB; W99022.
PT   Transgenic nematode containing nucleic acid related to
PT   triplet-repeat neurological disease - useful in screening for agents
PT   that can treat or prevent such diseases, especially Huntington's
PT   disease
PS   Claim 3; Fig 2; 40pp; English.
CC   The present invention describes a transgenic nematode (A) comprising a
CC   nucleic acid (I), associated with a triplet-repeat neurological disease
CC   (TREND), linked to a neuron-specific regulator. Also described are: (1)
CC   transgenic nematode (A') exhibiting progressive neuronal cell death; (2)
CC   expressing a nucleic acid in neuron-specific manner by linking it to an
CC   osm-10 regulator, in a region that, in the natural osm-10 gene, is
CC   occupied by all or part of the osm-10 coding sequence; and (3) transgene
CC   comprising (I) linked to the osm-10 regulator as specified in (2). (A)
CC   and (A') are used to screen for compounds useful in treatment or
CC   prevention of TRENDs (e.g. Huntington's disease, spinocerebellar ataxia,
CC   dentatorubral-pallidoluysian atrophy and spinal bulbar muscular
CC   dystrophy, also non-CAG related disorders). The nematodes are accurate
CC   models of TRENDs, i.e. they show the progressive nerve cell dysfunction
CC   characteristic of human disease and this can be monitored over time,
CC   using a population of animals that hatched together. Since they begin to
CC   show neuronal cell loss as early as day 7 from hatching, the nematodes
CC   can be used to test palliative treatments and the aetiology of TRENDs
CC   may also be studied. Nematodes have short generation times; large
CC   numbers can be manipulated easily, allowing accurate staging of many of
CC   them for high throughput testing, and can be bred to produce identical
CC   offspring. The osm-10 promoter is active only in neuronal tissue. The
CC   present sequence represents the human huntingtin gene from the present
CC   invention.
SQ   Sequence   515 BP;    125 A;    169 C;    133 G;    88 T;


  Query Match             1.1%;  Score 38.2;  DB 1;  Length 515;
  Best Local Similarity   51.6%;  Pred. No. 0.15;
  Matches  111;  Conservative    0;  Mismatches  103;  Indels    1;  Gaps    1;

Qy     74 GCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTG 133
          ||   ||||| || |      | |  ||||||  |  |||   ||| || | ||   |
Db    239 GCCGGGTGGCGGCGGGGGCGGCGGCGGGGGCGGCTGCGGCTGAGGCAGCAGCGGCTGTGC 180

Qy    134 CGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCGTTTAT-TCAGCAG 192
          | |    || ||| || |||  || | ||||||||| |||  | | ||    |  | || |
Db    179 CTGCGGCGGCGGCTGAGGAAGCTGAGGAGGCGGCGGCGGCGGCGGCGGTGGCGGCTG 120

Qy    193 GCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTGATTGGCAAGGCG 252
            |||    | | |  ||  ||   |||||  || | |  ||  |||| |      |  |
Db    119 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 60

Qy    253 ATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAA 287
          |    |||  || || ||| | |  | |  | |
Db     59 CTGCTGGAAGGACTTGAGGGACTCGAAGGCCTTCA 25


RESULT  14
Q31795/c
ID   Q31795 standard; cDNA; 342 BP.
AC   Q31795;
DT   21-APR-1993  (first entry)
DE   PRP2 gene.
KW   Proline-rich protein; PRP1; PRP2; PRP3; PRP4; PRP3g12; pistil;
KW   flower; sterile; plant; hydroxyproline; glycoprotein; gum; ss.
OS   Nicotiana alata.
FH   Key             Location/Qualifiers
FT   cds             1. .342
FT                   /*tag=  a
FT                   /label= PRP2
FT   repeat_region   58. .210
FT                   /*tag=  b
FT                   /note= "see CC"
FT   repeat_region   220. .327
FT                   /*tag=  c
FT                   /note= "see CC"
PN   WO9220713-A.
PD   26-NOV-1992.
PF   15-MAY-1992; AU0222.
PR   15-MAY-1991; AU-006159.
PA   (UYME ) UNIV MELBOURNE.
PI   Chen C, Clarke AE;
DR   WPI; 92-415710/50.
DR   P-PSDB; R29163.
PT   Isolated DNA sequence encoding proline rich proteins of Nicotiana
PT   alata - used to express useful proteins and identify regulatory
PT   elements for pistil specific expression, to give e.g. longer
PT   flower life
PS   Claim 2; Pag 47 + Fig 2; 57pp; English.
CC   The sequence below is the 342 bp insert from clone PRP2.1.
CC   The incomplete polypeptide encoded by this cDNA contains 49% proline
CC   residues in total of 114 amino acids.  It can be divided into two
CC   domains, one (encoded by nucleotides 58-210) contains 5 repeats of
CC   (Pro)4-Ala interspersed with 4 repeats of the pentapeptide Gln-Leu-
CC   Pro-Ile-Arg.  These repeat sequences in turn make up two copies
CC   of a higher order repetitive unit of 23 amino acids (Pro)4-Ala-Gln-
CC   Leu-Pro-Ile-Arg-Gln-(Pro)4-Ala-Thr-Gln-Leu-Pro-Ile-Arg-Lys.
CC   The other domain (nucleotides 220-327) is composed mainly of
CC   (Pro)2-5-Gly-Tyr repeats (7) that are tandemly reiterated.  At the
CC   nucleotide level, the sequences encoding the various repeats are
CC   highly conserved.
CC   The sequences of the PRP1, PRP2, PRP3 (cDNA and genomic DNA),
CC   PRP4 genes are given in Q31794-98.  The DNA sequences may be used to
CC   identify regulatory elements responsible for pistil-specific expression.
CC   Such elements may be used to provide pistil-specific expression of
CC   foreign or heterologous genes, e.g. to enable longer flower life or
CC   in the creation of male- or female-sterile plants.  The PRPs may be
CC   converted by hydroxylation and glycosylation into hydroxyproline-rich
CC   glycoproteins.  These could by used as effective substitutes for gums,
CC   e.g. guar gum and gum arabic, in foods and for non-food applications,
CC   where emulsification, thickening and stabilisation are required.
SQ   Sequence   342 BP;    105 A;    139 C;    53 G;    45 T;


  Query Match             1.1%;  Score 38;  DB 1;  Length 342;
  Best Local Similarity   45.4%;  Pred. No. 0.14;
  Matches  137;  Conservative    0;  Mismatches  165;  Indels    0;  Gaps    0;

Qy    123 GGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCGTT 182
          || ||    | |  |   |||||  |  || |   ||  |||||  |     ||| ||
Db    305 GGCGGCGGATACCCTGGCGGTGGTGGTGGATACCCTGGTGGCGGATATCCTGGTGGCGGA 246

Qy    183 TATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTGAT 242
          ||| | |  ||||   | |||| |  | ||| || |||||  |  | |  |     |  |
```

```
Db    245 TATCCTGGTGGCGGATATCCCGGTGGTTGAGTATATGCTGGTGGTGGTGGTTTCCTTATT 186

Qy    243 TGGCAAGGCGATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAGAT 302

Db    185 GGTAGCTGAGTAGCTGGTGGTGGTGGTTGCCTTATTGGTAACTGAGCTGGTGGTGGTGGT 126

Qy    303 GGCCGGCAGCATTGTTGGTGCGATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTGGT 362

Db    125 TTCCTTATTGGTAGCTGAGTAGCTGGTGGTGGTGGTTGCCTTATTGGTAACTGAGCTGGT 66

Qy    363 GGTCGCAGTTGTCGGGAAAGGCGCGGCGGCGAAACTGGGTAACGCGCTGAGCAAAATGAT 422

Db     65 GGTGGTGGTAGTGGTAATTTTGGTGGTGGTGCTCGTTGCATTGGCGGTGGCTGAGAAGGT 6

Qy    423 GG 424

Db      5 GG 4
```

RESULT 15
V20700/c
ID V20700 standard; DNA; 5163 BP.
AC V20700;
DT 17-AUG-1998 (first entry)
DE Cryptosporidium parvum GP900 antigen open reading frame.
KW Open reading frame; ORF; antigen; GP900; cryptosporidium; infection;
KW antibody; prophylaxis; treatment; inhibition; retardation;
KW detection; diagnosis; human; ds.
OS Cryptosporidium parvum.
FH Key Location/Qualifiers
FT CDS 1. .5163
FT /*tag= a
FT /product= GP900 antigen
FT misc_feature 524. .1270
FT /*tag= b
FT /note= "region containing NINC mutations"
PN WO9806430-A1.
PD 19-FEB-1998.
PF 11-AUG-1997; U14104.
PR 14-AUG-1996; US-700651.
PA (REGC) UNIV CALIFORNIA.
PI Gut J, Leech J, Nelson RC, Petersen C;
DR WPI; 98-159290/14.
DR P-PSDB; W48299.
PT Anti-Cryptosporidium antibody - used to develop products for
PT detection, diagnosis, prophylaxis or treatment of Cryptosporidium
PT infections
PS Claim 32; Pages 60-62; 89pp; English.
CC The sequence is that encoding the GP900 antigen which may be used
CC in the production of anti-Cryptosporidium antibodies. These can be
CC used for the prophylaxis, treatment, inhibition or retardation of
CC a Cryptosporidium infection in humans or in animals such as calves.
CC They can also be used for the detection and diagnosis of related
CC infections.
SQ Sequence 5163 BP; 1875 A; 1137 C; 873 G; 1278 T;

Query Match 1.1%; Score 37.8; DB 1; Length 5163;
Best Local Similarity 44.9%; Pred. No. 0.86;
Matches 144; Conservative 0; Mismatches 177; Indels 0; Gaps 0;

```
Qy     68 TTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCTGCGGTGG 127

Db    980 TTGTGGTGGTTGTAGTAGTAGTTGTTGTTGTTGTTGTTGTTGTAGTTGTTGTTGTTG 921

Qy    128 GACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCGTTTATTC 187

Db    920 GTTTCTTGGTAGTTGTGGTTGTGGTTGTGGTTGTTGTGGTTGTAGTTGTTGTGGTTGTAG 861

Qy    188 AGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTGATTGGCA 247

Db    860 TTGTGGTTGTAGTTGTTGTGGTTGTTGTTGTTGTGGTTGTAGTTGTTGTGGTTGTAGTTG 801

Qy    248 AGGCGATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAGATGGCCG 307

Db    800 TGGTTGTAGTTGTTGTGGTTGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAG 741

Qy    308 GCAGCATTGTTGGTGCGATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTGGTGGTCG 367

Db    740 TAGTCGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 681

Qy    368 CAGTTGTCGGGAAAGGCGCGG 388

Db    680 TTGTTGTCGTAGTAGTAGTGG 660
```

Search completed: October 5, 1999, 18:01:48
Job time: 11550 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 1999, 18:05:07 ; Search time 2414.69 Seconds
(without alignments)
4532.029 Million cell updates/sec

Title: US-09-068-804-15
Perfect score: 3441
Sequence: 1 CGCAAAGCCGAGGAAACGAA.....AAACAGGCGAGTCTGGATGC 3441

Scoring table: IDENTITY_NUC

Searched: 679419 seqs, 1590154680 residues

Database :
GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_st:*
14: gb_sts:*
15: gb_sy:*
16: gb_un:*
17: gb_vi:*
18: em_fun:*
19: em_htg:*
20: em_hum1:*
21: em_hum2:*
22: em_in:*
23: em_om:*
24: em_or:*
25: em_ov:*
26: em_pat:*
27: em_ph:*
28: em_pl:*
29: em_ro:*
30: em_sts:*
31: em_sy:*
32: em_un:*
33: em_vi:*
34: gb_htg1:*
35: gb_htg2:*
36: gb_in1:*
37: gb_in2:*
38: em_ba1:*
39: em_ba2:*
40: em_hum3:*
41: em_hum4:*
42: gb_pr4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3441	100.0	3441	2	STU30491	U30491 Salmonella
2	3179.8	92.4	5393	1	STSIPGNS	X82670 S.typhi spa
3	2063	60.0	3622	1	STU25631	U25631 Salmonella
4	1443.4	41.9	2995	1	SDU66877	U66877 Salmonella
5	1290.4	37.5	3500	1	STU40013	U40013 Salmonella
6	179	5.2	6869	1	SDIPABCD	X60777 Shigella dy
7	179	5.2	7663	1	SFPVR2	X15319 Shigella fl
8	179	5.2	5255	1	SHFIPABCD	J04117 S.flexneri
9	179	5.2	4761	1	W10PPIBCA	M34849 Plasmid pWR
10	72.6	2.1	207466	35	AC007249	AC007249 Homo sapi
11	72	2.1	158880	34	AC007009	AC007009 Homo sapi
c 12	71.4	2.1	137137	42	AC004928	AC004928 Homo sapi
13	70.2	2.0	168844	35	AC007263	AC007263 Homo sapi
c 14	67.8	2.0	162269	11	AC000097	AC000097 Homo sapi
c 15	67.8	2.0	165321	35	AC006549	AC006549 Homo sapi
c 16	67.8	2.0	179269	42	AC006547	AC006547 , complet
17	67.6	2.0	171752	35	AC007381	AC007381 Homo sapi
c 18	67.4	2.0	141077	35	AC005918	AC005918 Homo sapi
19	65.4	1.9	858	10	HSHZREP42	X59828 Human chrom
20	62.8	1.8	103826	34	HSB11L7C8	AL078472 Homo sapi
c 21	61.6	1.8	41185	10	HSB1L1C6	AL049849 Homo sapi
22	59.6	1.7	131453	34	HS989H11	Z83851 Homo sapien
23	58.4	1.7	256172	35	AC005139	AC005139 Plasmodiu
c 24	57.4	1.7	1988	11	AC002494	AC002494 Human PCR
25	56.8	1.7	27251	9	HS293L6C	Z83732 Human DNA s
26	56.8	1.7	99497	10	HS293L6	AL049749 Human DNA
c 27	56.6	1.6	7218	5	I66494	I66494 Sequence 14
c 28	55.8	1.6	175484	34	HS215D11	AL034417 Homo sapi
c 29	55.2	1.6	284972	35	AC005505	AC005505 Plasmodiu
c 30	55	1.6	143701	9	HS29C18	Z97192 Homo sapien
31	54.6	1.6	174428	11	AC005821	AC005821 Homo sapi
c 32	54.2	1.6	131427	9	HS37J18	Z98747 Human DNA s
c 33	54.2	1.6	175609	11	AC005243	AC005243 Homo sapi
34	54	1.6	112460	34	HS1104E15	AL022312 Homo sapi
35	53	1.5	155074	11	AC003693	AC003693 Human Chr
36	53	1.5	105892	34	HSY214H10_4	Continuation (5 of
c 37	52.6	1.5	135258	9	HS127B20	Z83838 Human DNA s
38	51.4	1.5	834	7	OSGRP09MR	X54449 O.sativa GR
39	51.4	1.5	3250	7	OSU40708	U40708 Oryza sativ
c 40	51.2	1.5	117338	10	HS173D1	AL031984 Human DNA
41	50.2	1.5	888	7	NTGRPR	X74106 N.tabacum g
42	49.8	1.4	1925	7	PVGRP18	X13596 Bean DNA fo
43	49.8	1.4	206344	35	AC007240	AC007240 Homo sapi
c 44	49.8	1.4	242867	35	AC007623	AC007623 Homo sapi
45	49	1.4	103270	7	ATF1N20	AL022140 Arabidops

ALIGNMENTS

RESULT 1
STU30491
LOCUS STU30491 3441 bp DNA BCT 19-JAN-1996
DEFINITION Salmonella typhimurium SspA (sspA) and SspB (sspB) genes, partial cds, and SspC (sspC) and SspD (sspD) genes, complete cds.
ACCESSION U30491
NID g1117811
VERSION U30491.1 GI:1117811
KEYWORDS .
SOURCE Salmonella typhimurium.
ORGANISM Salmonella typhimurium
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Salmonella.
REFERENCE 1 (bases 1 to 3441)
AUTHORS Hueck,C.J., Hantman,M.J., Bajaj,V., Johnston,C., Lee,C.A. and Miller,S.I.
TITLE Salmonella typhimurium secreted invasion determinants are homologous to Shigella Ipa proteins
JOURNAL Mol. Microbiol. 18 (3), 479-490 (1995)
MEDLINE 96342378
REFERENCE 2 (bases 1 to 3441)
AUTHORS Miller,S.I.
TITLE Direct Submission
JOURNAL Submitted (28-JUN-1995) Samuel I. Miller, Infectious Disease Unit,

```
          Massachusetts General Hospital, Harvard Medical School, Fruit
          Street, Boston, MA 02114, USA
FEATURES             Location/Qualifiers
     source          1. .3441
                     /organism="Salmonella typhimurium"
                     /strain="SL1344"
                     /db_xref="taxon:602"
                     /map="59 min"
     gene            1. .870
                     /gene="sspB"
     CDS             <1. .870
                     /gene="sspB"
                     /note="secreted protein"
                     /codon_start=1
                     /transl_table=11
                     /function="invasion determinant"
                     /product="SspB"
                     /protein_id="AAC43546.1"
                     /db_xref="PID:g1117812"
                     /db_xref="GI:1117812"
                     /translation="RKAEETNRIMGCIGKVLGALLTIVSVVAAVFTGGASLALAAVGL
                     AVMVADEIVKAATGVSFIQQALNPIMEHVLKPLMELIGKAITKALEGLGVDKKTAEMA
                     GSIVGAIVAAIAMVAVIVVVAVVGKGAAAKLGNALSKMMGETIKKLVPNVLKQLAQNG
                     SKLFTQGMQRITSGLGNVGSKMGLQTNALSKELVGNTLNKVALGMEVTNTAAQSAGGV
                     AEGVFIKNASEALADFMLARFAMDQIQQWLKQSVEIFGENQKVTAELQKAMSSAVQQN
                     ADASRFILRQSRA"
     gene            898. .2127
                     /gene="sspC"
     CDS             898. .2127
                     /gene="sspC"
                     /note="secreted protein"
                     /codon_start=1
                     /transl_table=11
                     /function="invasion determinant"
                     /evidence=experimental
                     /product="SspC"
                     /protein_id="AAC43547.1"
                     /db_xref="PID:g1117813"
                     /db_xref="GI:1117813"
                     /translation="MLISNVGINPAAYLNNHSVENSSQTASQSVSAKDILNSIGISSS
                     KVSDLGLSPTLSAPAPGVLTQTPGTITSFLKASIQNTDMNQDLNALANNVTTKANEVV
                     QTQLREQQAEVGKFFDISGMSSSAVALLAAANTLMLTLNQADSKLSGKLSLVSFDAAK
                     TTASSMMREGMNALSGSISQSALQLGITGVGAKLEYKGLQNERGALKHNAAKIDKLTT
                     ESHSIKNVLNGQNSVKLGAEGVDSLKSLNMKKTGTDATKNLNDATLKSNAGTSATESL
                     GIKDSNKQISPEHQAILSKRLESVESDIRLEQNTMDMTRIDARKMQMTGDLIMKNSVT
                     VGGIAGASGQYAATQERSEQQISQVNNRVASTASDEARESSRKSTSLIQEMLKTMESI
                     NQSKASALAAIAGNIRA"
     gene            2198. .3229
                     /gene="sspD"
     CDS             2198. .3229
                     /gene="sspD"
                     /note="secreted protein"
                     /codon_start=1
                     /transl_table=11
                     /function="invasion determinant"
                     /evidence=experimental
                     /product="SspD"
                     /protein_id="AAC43548.1"
                     /db_xref="PID:g1117814"
                     /db_xref="GI:1117814"
                     /translation="MLNIQNYSASPHPGIVAERPQTPSASEHVETAVVPSTTEHRGTD
                     IISLSQAATKIHQAQQTLQSTPPISEENNDERTLARQQLTSSLNALAKSGVSLSAEQN
                     ENLRSAFSAPTSALFSASPMAQPRTTISDAEIWDMVSQNISAIGDSYLGVYENVVAVY
                     TDFYQAFSDILSKMGGWLLPGKDGNTVKLDVTSLKNDLNSLVNKYNQINSNTVLFPAQ
                     SGSGVKVATEAEARQWLSELNLPNSCLKSYGSGYVVTVDLTPLQKMVQDIDGLGAPGK
                     DSKLEMDNAKYQAWQSGFKAQEENMKTTLQTLTQKYSNANSLYDNLVKVLSSTISSSL
                     ETAKSFLQG"
     gene            3248. .3441
                     /gene="sspA"
     CDS             3248. .>3441
                     /gene="sspA"
                     /note="secreted protein"
                     /codon_start=1
                     /transl_table=11
                     /function="unknown"
                     /product="SspA"
                     /protein_id="AAC43549.1"
                     /db_xref="PID:g1117815"
                     /db_xref="GI:1117815"
                     /translation="MVTSVRTQPPVIMPGMQTEIKTQATNLAANLSAVRESATATLSG
                     EIKGPQLEDFPALIKQASLDA"
BASE COUNT       978 a    759 c    925 g    779 t
ORIGIN
```

```
  Query Match           100.0%;  Score 3441;  DB 2;  Length 3441;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 3441;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCAAAGCCGAGGAAACGAACCGCATTATGGGATGTATCGGGAAAGTCCTCGGCGCGCTG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CGCAAAGCCGAGGAAACGAACCGCATTATGGGATGTATCGGGAAAGTCCTCGGCGCGCTG 60

Qy      61 CTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCT 120

Qy     121 GCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCG 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCG 180

Qy     181 TTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTG 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTG 240

Qy     241 ATTGGCAAGGCGATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAG 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 ATTGGCAAGGCGATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAG 300

Qy     301 ATGGCCGGCAGCATTGTTGGTGCGATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTG 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 ATGGCCGGCAGCATTGTTGGTGCGATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTG 360

Qy     361 GTGGTCGCAGTTGTCGGGAAAGGCGCGGCGGCGAAACTGGGTAACGCGCTGAGCAAAATG 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GTGGTCGCAGTTGTCGGGAAAGGCGCGGCGGCGAAACTGGGTAACGCGCTGAGCAAAATG 420

Qy     421 ATGGGCGAAACGATTAAGAAGTTGGTGCCTAACGTGCTGAAACAGTTGGCGCAAAACGGC 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 ATGGGCGAAACGATTAAGAAGTTGGTGCCTAACGTGCTGAAACAGTTGGCGCAAAACGGC 480

Qy     481 AGCAAACTCTTTACCCAGGGGATGCAACGTATTACTAGCGGTCTGGGTAATGTGGGTAGC 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 AGCAAACTCTTTACCCAGGGGATGCAACGTATTACTAGCGGTCTGGGTAATGTGGGTAGC 540

Qy     541 AAGATGGGCCTGCAAACGAATGCCTTAAGTAAAGAGCTGGTAGGTAATACCCTAAATAAA 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 AAGATGGGCCTGCAAACGAATGCCTTAAGTAAAGAGCTGGTAGGTAATACCCTAAATAAA 600

Qy     601 GTGGCGTTGGGCATGGAAGTCACGAATACCGCAGCCCAGTCAGCCGGTGGTGTTGCCGAG 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 GTGGCGTTGGGCATGGAAGTCACGAATACCGCAGCCCAGTCAGCCGGTGGTGTTGCCGAG 660

Qy     661 GGCGTATTTATTAAAAATGCCAGCGAGGCGCTTGCTGATTTTATGCTCGCCCGTTTTGCC 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 GGCGTATTTATTAAAAATGCCAGCGAGGCGCTTGCTGATTTTATGCTCGCCCGTTTTGCC 720

Qy     721 ATGGATCAGATTCAGCAGTGGCTTAAACAATCCGTAGAAATATTTGGTGAAAACCAGAAG 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 ATGGATCAGATTCAGCAGTGGCTTAAACAATCCGTAGAAATATTTGGTGAAAACCAGAAG 780

Qy     781 GTAACGGCGGAACTGCAAAAAGCCATGTCTTCTGCGGTACAGCAAAATGCGGATGCTTCG 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 GTAACGGCGGAACTGCAAAAAGCCATGTCTTCTGCGGTACAGCAAAATGCGGATGCTTCG 840

Qy     841 CGTTTTATTCTGCGCCAGAGTCGCGCATAAAAACTGCCAAAATAAAGGGAGAAAAATATG 900
```

```
Db    841 CGTTTTATTCTGCGCCAGAGTCGCGCATAAAAACTGCCAAAATAAAGGGAGAAAAATATG 900
Qy    901 TTAATTAGTAATGTGGGAATAAATCCCGCCGCTTATTTAAATAATCATTCTGTTGAGAAT 960
Db    901 TTAATTAGTAATGTGGGAATAAATCCCGCCGCTTATTTAAATAATCATTCTGTTGAGAAT 960
Qy    961 AGTTCACAGACAGCTTCGCAATCCGTTAGCGCTAAAGATATTCTGAATAGTATTGGTATT 1020
Db    961 AGTTCACAGACAGCTTCGCAATCCGTTAGCGCTAAAGATATTCTGAATAGTATTGGTATT 1020
Qy   1021 AGCAGCAGTAAAGTCAGTGACCTGGGGTTGAGTCCTACACTGAGCGCGCCTGCGCCAGGG 1080
Db   1021 AGCAGCAGTAAAGTCAGTGACCTGGGGTTGAGTCCTACACTGAGCGCGCCTGCGCCAGGG 1080
Qy   1081 GTATTAACGCAAACCCCCGGAACGATCACGTCCTTTTTAAAAGCCAGTATTCAAAATACC 1140
Db   1081 GTATTAACGCAAACCCCCGGAACGATCACGTCCTTTTTAAAAGCCAGTATTCAAAATACC 1140
Qy   1141 GACATGAATCAGGATTTGAATGCTCTGGCAAATAATGTCACGACTAAAGCGAATGAGGTT 1200
Db   1141 GACATGAATCAGGATTTGAATGCTCTGGCAAATAATGTCACGACTAAAGCGAATGAGGTT 1200
Qy   1201 GTGCAAACCCAGTTACGCGAGCAGCAGGCAGAAGTCGGAAAGTTTTTTGATATTAGCGGA 1260
Db   1201 GTGCAAACCCAGTTACGCGAGCAGCAGGCAGAAGTCGGAAAGTTTTTTGATATTAGCGGA 1260
Qy   1261 ATGTCTTCCAGTGCCGTTGCGCTGTTGGCTGCCGCGAATACGTTAATGCTGACGTTGAAC 1320
Db   1261 ATGTCTTCCAGTGCCGTTGCGCTGTTGGCTGCCGCGAATACGTTAATGCTGACGTTGAAC 1320
Qy   1321 CAGGCTGATAGCAAACTGTCTGGTAAGTTGTCATTAGTCAGTTTTGATGCAGCTAAAACG 1380
Db   1321 CAGGCTGATAGCAAACTGTCTGGTAAGTTGTCATTAGTCAGTTTTGATGCAGCTAAAACG 1380
Qy   1381 ACGGCAAGCTCCATGATGCGCGAAGGGATGAATGCGTTGTCCGGTAGTATTTCCCAGAGC 1440
Db   1381 ACGGCAAGCTCCATGATGCGCGAAGGGATGAATGCGTTGTCCGGTAGTATTTCCCAGAGC 1440
Qy   1441 GCGCTTCAGTTGGGGATCACTGGCGTGGGCGCCAAACTGGAATATAAGGGGCTGCAGAAT 1500
Db   1441 GCGCTTCAGTTGGGGATCACTGGCGTGGGCGCCAAACTGGAATATAAGGGGCTGCAGAAT 1500
Qy   1501 GAAAGAGGCGCGCTTAAACATAATGCCGCGAAGATCGATAAACTGACCACTGAAAGCCAC 1560
Db   1501 GAAAGAGGCGCGCTTAAACATAATGCCGCGAAGATCGATAAACTGACCACTGAAAGCCAC 1560
Qy   1561 AGTATTAAAAACGTGCTGAACGGGCAGAATAGCGTCAAACTCGGTGCTGAAGGCGTCGAT 1620
Db   1561 AGTATTAAAAACGTGCTGAACGGGCAGAATAGCGTCAAACTCGGTGCTGAAGGCGTCGAT 1620
Qy   1621 TCTCTGAAATCGTTAAATATGAAGAAAACCGGTACCGATGCGACGAAAAATCTTAATGAT 1680
Db   1621 TCTCTGAAATCGTTAAATATGAAGAAAACCGGTACCGATGCGACGAAAAATCTTAATGAT 1680
Qy   1681 GCGACGCTTAAATCTAATGCCGGAACCAGCGCCACGGAAAGTCTGGGTATTAAAGACAGT 1740
Db   1681 GCGACGCTTAAATCTAATGCCGGAACCAGCGCCACGGAAAGTCTGGGTATTAAAGACAGT 1740
Qy   1741 AATAAACAAATCTCCCCTGAACATCAGGCTATTCTGTCGAAACGTCTTGAGTCTGTCGAA 1800
Db   1741 AATAAACAAATCTCCCCTGAACATCAGGCTATTCTGTCGAAACGTCTTGAGTCTGTCGAA 1800
Qy   1801 TCCGATATTCGTCTTGAGCAGAATACCATGGATATGACCCGAATCGATGCGCGCAAGATG 1860
Db   1801 TCCGATATTCGTCTTGAGCAGAATACCATGGATATGACCCGAATCGATGCGCGCAAGATG 1860
Qy   1861 CAGATGACGGGCGATCTGATTATGAAGAACTCGGTCACGGTCGGTGGTATTGCAGGGGCG 1920
Db   1861 CAGATGACGGGCGATCTGATTATGAAGAACTCGGTCACGGTCGGTGGTATTGCAGGGGCG 1920
Qy   1921 TCCGGGCAGTACGCCGCTACTCAGGAACGTTCCGAGCAGCAAATTAGCCAGGTGAATAAC 1980
Db   1921 TCCGGGCAGTACGCCGCTACTCAGGAACGTTCCGAGCAGCAAATTAGCCAGGTGAATAAC 1980
Qy   1981 CGGGTTGCCAGCACCGCATCGGACGAAGCCCGTGAAAGTTCACGTAAATCGACCAGCCTG 2040
Db   1981 CGGGTTGCCAGCACCGCATCGGACGAAGCCCGTGAAAGTTCACGTAAATCGACCAGCCTG 2040
Qy   2041 ATTCAGGAAATGCTGAAAACAATGGAGAGCATTAACCAGTCGAAAGCATCCGCACTCGCT 2100
Db   2041 ATTCAGGAAATGCTGAAAACAATGGAGAGCATTAACCAGTCGAAAGCATCCGCACTCGCT 2100
Qy   2101 GCTATCGCAGGCAATATTCGCGCTTAATCTGAAAGGTCATCTATACGCCATCATGGGTGT 2160
Db   2101 GCTATCGCAGGCAATATTCGCGCTTAATCTGAAAGGTCATCTATACGCCATCATGGGTGT 2160
Qy   2161 GATTTAATCGCGCTCCTGATGGCGAACTGGGGATATTATGCTTAATATTCAAAATTATTC 2220
Db   2161 GATTTAATCGCGCTCCTGATGGCGAACTGGGGATATTATGCTTAATATTCAAAATTATTC 2220
Qy   2221 CGCTTCTCCTCATCCGGGGATCGTTGCCGAACGGCCGCAGACTCCCTCGGCGAGCGAGCA 2280
Db   2221 CGCTTCTCCTCATCCGGGGATCGTTGCCGAACGGCCGCAGACTCCCTCGGCGAGCGAGCA 2280
Qy   2281 CGTCGAGACTGCCGTGGTACCGTCTACCACAGAACATCGCGGTACAGATATCATTTCATT 2340
Db   2281 CGTCGAGACTGCCGTGGTACCGTCTACCACAGAACATCGCGGTACAGATATCATTTCATT 2340
Qy   2341 ATCGCAGGCGGCTACTAAAATCCACCAGGCACAGCAGACGCTGCAGTCAACGCCACCGAT 2400
Db   2341 ATCGCAGGCGGCTACTAAAATCCACCAGGCACAGCAGACGCTGCAGTCAACGCCACCGAT 2400
Qy   2401 CTCTGAAGAGAATAATGACGAGCGCACGCTGGCGCGCCAGCAGTTGACCAGCAGCCTGAA 2460
Db   2401 CTCTGAAGAGAATAATGACGAGCGCACGCTGGCGCGCCAGCAGTTGACCAGCAGCCTGAA 2460
Qy   2461 TGCGCTGGCGAAGTCCGGCGTGTCATTATCCGCAGAACAAAATGAGAACCTGCGGAGCGC 2520
Db   2461 TGCGCTGGCGAAGTCCGGCGTGTCATTATCCGCAGAACAAAATGAGAACCTGCGGAGCGC 2520
Qy   2521 GTTTTCTGCGCCGACGTCGGCCTTATTTAGCGCTTCGCCTATGGCGCAGCCGAGAACAAC 2580
Db   2521 GTTTTCTGCGCCGACGTCGGCCTTATTTAGCGCTTCGCCTATGGCGCAGCCGAGAACAAC 2580
Qy   2581 CATTTCTGATGCTGAGATTTGGGATATGGTTTCCCAAAATATATCGGCGATAGGTGACAG 2640
Db   2581 CATTTCTGATGCTGAGATTTGGGATATGGTTTCCCAAAATATATCGGCGATAGGTGACAG 2640
Qy   2641 CTATCTGGGCGTTTATGAAAACGTTGTCGCAGTCTATACCGATTTTTATCAGGCCTTCAG 2700
Db   2641 CTATCTGGGCGTTTATGAAAACGTTGTCGCAGTCTATACCGATTTTTATCAGGCCTTCAG 2700
Qy   2701 TGATATTCTTTCCAAAATGGGAGGCTGGTTATTACCAGGTAAGGACGGTAATACCGTTAA 2760
Db   2701 TGATATTCTTTCCAAAATGGGAGGCTGGTTATTACCAGGTAAGGACGGTAATACCGTTAA 2760
Qy   2761 GCTAGATGTTACCTCACTCAAAAATGATTTAAACAGTTTAGTCAATAAATATAATCAAAT 2820
Db   2761 GCTAGATGTTACCTCACTCAAAAATGATTTAAACAGTTTAGTCAATAAATATAATCAAAT 2820
Qy   2821 AAACAGTAATACCGTTTTATTTCCAGCGCAGTCAGGCAGCGGCGTTAAAGTAGCCACTGA 2880
Db   2821 AAACAGTAATACCGTTTTATTTCCAGCGCAGTCAGGCAGCGGCGTTAAAGTAGCCACTGA 2880
Qy   2881 AGCGGAAGCGAGACAGTGGCTCAGTGAATTGAATTTACCGAATAGCTGCCTGAAATCTTA 2940
Db   2881 AGCGGAAGCGAGACAGTGGCTCAGTGAATTGAATTTACCGAATAGCTGCCTGAAATCTTA 2940
Qy   2941 TGGATCCGGTTATGTCGTCACCGTTGATCTGACGCCATTACAAAAAATGGTTCAGGATAT 3000
Db   2941 TGGATCCGGTTATGTCGTCACCGTTGATCTGACGCCATTACAAAAAATGGTTCAGGATAT 3000
Qy   3001 TGATGGTTTAGGCGCGCCGGGAAAAGACTCAAAACTCGAAATGGATAACGCCAAATATCA 3060
Db   3001 TGATGGTTTAGGCGCGCCGGGAAAAGACTCAAAACTCGAAATGGATAACGCCAAATATCA 3060
```

```
Qy   3061 AGCCTGGCAGTCGGGTTTTAAAGCGCAGGAAGAAAATATGAAAACCACATTACAGACGCT 3120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3061 AGCCTGGCAGTCGGGTTTTAAAGCGCAGGAAGAAAATATGAAAACCACATTACAGACGCT 3120

Qy   3121 GACGCAAAAATATAGCAATGCCAATTCATTGTACGACAACCTGGTAAAAGTGCTGAGCAG 3180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3121 GACGCAAAAATATAGCAATGCCAATTCATTGTACGACAACCTGGTAAAAGTGCTGAGCAG 3180

Qy   3181 TACGATAAGTAGCAGCCTGGAAACCGCCAAAAGCTTCCTGCAAGGATAACAGAAGAGGAT 3240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3181 TACGATAAGTAGCAGCCTGGAAACCGCCAAAAGCTTCCTGCAAGGATAACAGAAGAGGAT 3240

Qy   3241 ATTAATAATGGTTACAAGTGTAAGGACTCAGCCCCCCGTCATAATGCCAGGTATGCAGAC 3300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3241 ATTAATAATGGTTACAAGTGTAAGGACTCAGCCCCCCGTCATAATGCCAGGTATGCAGAC 3300

Qy   3301 CGAGATCAAAACGCAGGCCACGAATCTTGCGGCGAATCTTTCCGCAGTCAGAGAAAGTGC 3360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3301 CGAGATCAAAACGCAGGCCACGAATCTTGCGGCGAATCTTTCCGCAGTCAGAGAAAGTGC 3360

Qy   3361 CACAGCGACGCTGTCAGGGGAAATTAAAGGCCCGCAACTGGAAGATTTTCCCGCGCTGAT 3420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3361 CACAGCGACGCTGTCAGGGGAAATTAAAGGCCCGCAACTGGAAGATTTTCCCGCGCTGAT 3420

Qy   3421 CAAACAGGCGAGTCTGGATGC 3441
          |||||||||||||||||||||
Db   3421 CAAACAGGCGAGTCTGGATGC 3441
```

```
RESULT   2
STSIPGNS
LOCUS       STSIPGNS     5393 bp    DNA             BCT       04-OCT-1995
DEFINITION  S.typhi spaT, sipB, sipC, sipD and sipA genes.
ACCESSION   X82670
NID         g1009001
VERSION     X82670.1  GI:1009001
KEYWORDS    sipA gene; sipB gene; sipC gene; sipD gene; spaT gene.
SOURCE      Salmonella typhi.
  ORGANISM  Salmonella typhi
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Salmonella.
REFERENCE   1  (bases 1 to 5393)
  AUTHORS   Hermant,D., Menard,R., Arricau,N., Parsot,C. and Popoff,M.Y.
  TITLE     Functional conservation of the Salmonella and Shigella effectors of
            entry into epithelial cells
  JOURNAL   Mol. Microbiol. 17 (4), 781-789 (1995)
  MEDLINE   96111497
REFERENCE   2  (bases 1 to 5393)
  AUTHORS   Popoff,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-1994) M. Popoff, Institut Pasteur, Unite des
            Enterobacteries, 28 rue du Docteur Roux, 75724 Paris Cedex 15,
            FRANCE
FEATURES             Location/Qualifiers
     source          1. .5393
                     /organism="Salmonella typhi"
                     /strain="Ty2"
                     /db_xref="taxon:601"
     gene            43. .540
                     /gene="spaT"
     CDS             43. .540
                     /gene="spaT"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA57987.1"
                     /db_xref="PID:g1009002"
                     /db_xref="GI:1009002"
                     /db_xref="SPTREMBL:Q57260"
                     /translation="MDYQNNVSEERVAEMIWDAVSEGATLKDVHGIPQDMMDGLYAHA
                     YEFYNQGRLDEAETFFRFLCIYDFYNPDYTMGLAAVCQLKKQFQKACDLYAVAFTLLK
                     NDYRPVFFTGQCQLLMRKAAKARQCFELVNERTEDESLRAKALVYLEALKTAETEQHS
                     EQEKE"
     gene            543. .2324
                     /gene="sipB"
     CDS             543. .2324
                     /gene="sipB"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA57988.1"
                     /db_xref="PID:g1009003"
                     /db_xref="GI:1009003"
                     /db_xref="SPTREMBL:Q56134"
                     /translation="MVNDASSISRSGYTQNPRLAEAAFEGVRKNTDFLKAADKAFKDV
                     VATKAGDLKAGTKSGESAINTVGLKPPTDAAREKLSSEGQLTLLLGKLMTLLGDVSLS
                     QLESRLAVWQAMIESQKEMGIQVSKEFQTALGEAQEATDLYEASIKKTDTAKSVYDAA
                     AKKLTQAQNKLQSLDPADPGYAQAEAAVEQAGKEATEAKEALDKATDATVKAGTDAKA
                     KAEKADNILTKFQGTANAASQNQVSQGEQDNLSNVARLTMLMAMFIEIVGKNTEESLQ
                     NDLALFNALQEGRQAEMEKKSAEFQEETRKAEETNRIMGCIGKVLGALLTIVSVVAAV
                     FTGGASLALAAVGLAVMVADEIVKAATGVSFIQQALNPIMEHVLKPLMELIGKAITKA
                     LEGLGVDKKTAEMAGSIVGAIVAAIAMVAVIVVVAVVGKGAAAKLGNALSKMMGETIK
                     KLVPNVLKQLAQNGSKLFTQGMQRITSGLGNVGSKMGLQTNALSKELVGNTLNKVALG
                     MEVTNTAAQSAGGVAEGVFIKNASEALADFMLARFAMDQIQQWLKQSVEIFGENQKVT
                     AELQKAMSSAVQQNADASRFILRQSRA"
     gene            2352. .3581
                     /gene="sipC"
     CDS             2352. .3581
                     /gene="sipC"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA57989.1"
                     /db_xref="PID:g1009004"
                     /db_xref="GI:1009004"
                     /db_xref="SPTREMBL:Q56135"
                     /translation="MLISNVGINPAAYLNNHSVENSSQTASQSVSAKDILNSIGISSS
                     KVSDLGLSPTLSAPAPGVLTQTPGTITSFLKASIQNTDMNQDLNALANNVTTKANEVV
                     QTQLREQQAEVGKFFDISGMSSSAVALLAAANTLMLTLNQADSKLSGKLSLVSFDAAK
                     TTASSMMREGMNALSGSISQSALQLGITGVGAKLEYKGLQNERGALKHNAAKIDKLTT
                     ESHSIKNVLNGQNSVKLGAEGVDSLKSLNMKKTGTDATKNLNDATLKSNAGTSATESL
                     GIKNSNKQISPEHQAILSKRLESVESDIRLEQNTMDMTRIDARKMQMTGDLIMKNSVT
                     VGGIAGASRQYAATQERSEQQISQVNNRVASTASDEARESSRKSTSLIQEMLKTMESI
                     NQSKASALAAIAGNIRA"
     gene            3651. .4670
                     /gene="sipD"
     CDS             3651. .4670
                     /gene="sipD"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA57990.1"
                     /db_xref="PID:g1009005"
                     /db_xref="GI:1009005"
                     /db_xref="SPTREMBL:Q56136"
                     /translation="MLNIQNYSASPHPGIVAERPQTPSASEHAEIAVVPSTTEHRGTD
                     IISLSQAATKIQQAQQTLQSTPPISEENNDERTLARQQLTSSLNALAKSGVSLSAEQN
                     ENLRSTFSARRRPYLALRLWPARTTISDAEIWDMVSQNISAIGDSYLGVYENVVAVYT
                     DFYQAFSDILSKMGGWLSPGKDGNTIKLNVDSLKSEISSLINKYTQINKNTILFPSQT
                     GSGMTTATKAEAEQWIKELNLPDSCLKASGSGYVVLVDTGPLSKMVSDLNGIGSGSAL
                     ELDNAKYQAWQSGFKAQEENLKTTLQTLTQKYSNANSLYDNLVKVLSSTISSSLETAK
                     SFLQG"
     gene            4689. .5393
                     /gene="sipA"
     CDS             4689. .>5393
                     /gene="sipA"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA57991.1"
                     /db_xref="PID:g1009006"
                     /db_xref="GI:1009006"
                     /db_xref="SPTREMBL:Q56137"
                     /translation="MVTSVRTQPPVIMPGMQTEIKTQATNLAANLSAVRESATATLSG
                     EIKGQQLEDFPALIKQASLDALFKCGKDAEALKEVFTNSNNVAGKKAIMEFAGLFRSA
                     LNATSDSPEAKTLLMKVGAEYTAQIIKDGLKEKSAFGPWLPETKKAEAKLENLEKQLL
                     DIIKNNTGGELSKLSTNLVMQEVMPYIASCIEHNFGCTLDPLTRSSLTQLVDKAAAKA
                     VEALDMCHQKLTQEQGT"
BASE COUNT     1584 a   1156 c   1458 g   1195 t
ORIGIN
```

Query Match 92.4%; Score 3179.8; DB 1; Length 5393;
Best Local Similarity 96.1%; Pred. No. 0;
Matches 3306; Conservative 0; Mismatches 122; Indels 13; Gaps 4;

```
Qy       1 CGCAAAGCCGAGGAAACGAACCGCATTATGGGATGTATCGGGAAAGTCCTCGGCGCGCTG 60
Db    1455 CGCAAAGCCGAGGAAACGAACCGCATTATGGGATGTATCGGGAAAGTCCTCGGCGCGCTG 1514

Qy      61 CTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCT 120
Db    1515 CTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCT 1574

Qy     121 GCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCG 180
Db    1575 GCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGGGTGTCG 1634

Qy     181 TTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTG 240
Db    1635 TTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTG 1694

Qy     241 ATTGGCAAGGCGATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAG 300
Db    1695 ATTGGCAAGGCGATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAG 1754

Qy     301 ATGGCCGGCAGCATTGTTGGTGCGATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTG 360
Db    1755 ATGGCAGGCAGCATTGTTGGTGCGATTGTCGCCGCTATTGCCATGGTAGCGGTCATTGTG 1814

Qy     361 GTGGTCGCAGTTGTCGGGAAAGGCGCGGCGGCGAAACTGGGTAACGCGCTGAGCAAAATG 420
Db    1815 GTGGTCGCAGTTGTCGGGAAAGGCGCGGCGGCGAAACTGGGTAACGCGCTGAGCAAAATG 1874

Qy     421 ATGGGCGAAACGATTAAGAAGTTGGTGCCTAACGTGCTGAAACAGTTGGCGCAAAACGGC 480
Db    1875 ATGGGCGAAACGATTAAGAAGTTGGTGCCTAACGTGCTGAAACAGTTGGCACAAAACGGC 1934

Qy     481 AGCAAACTCTTTACCCAGGGGATGCAACGTATTACTAGCGGTCTGGGTAATGTGGGTAGC 540
Db    1935 AGCAAACTCTTTACCCAGGGGATGCAACGTATTACTAGCGGCCTGGGTAATGTGGGTAGC 1994

Qy     541 AAGATGGGCCTGCAAACGAATGCCTTAAGTAAAGAGCTGGTAGGTAATACCCTAAATAAA 600
Db    1995 AAGATGGGCCTGCAAACGAATGCCTTAAGTAAAGAGCTGGTAGGTAATACCCTAAATAAA 2054

Qy     601 GTGGCGTTGGGCATGGAAGTCACGAATACCGCAGCCCAGTCAGCCGGTGGTGTTGCCGAG 660
Db    2055 GTGGCGTTGGGCATGGAAGTCACGAATACCGCAGCCCAGTCAGCCGGTGGGGTTGCCGAG 2114

Qy     661 GGCGTATTTATTAAAAATGCCAGCGAGGCGCTTGCTGATTTTATGCTCGCCCGTTTTGCC 720
Db    2115 GGGGTATTTATTAAAAATGCCAGCGAGGCGCTTGCTGATTTTATGCTCGCCCGTTTTGCC 2174

Qy     721 ATGGATCAGATTCAGCAGTGGCTTAAACAATCCGTAGAAATATTTGGTGAAAACCAGAAG 780
Db    2175 ATGGATCAGATTCAGCAGTGGCTTAAACAATCCGTAGAAATATTTGGTGAAAACCAGAAG 2234

Qy     781 GTAACGGCGGAACTGCAAAAAGCCATGTCTTCTGCGGTACAGCAAAATGCGGATGCTTCG 840
Db    2235 GTAACGGCGGAACTGCAAAAAGCCATGTCTTCTGCGGTACAGCAAAATGCGGATGCTTCG 2294

Qy     841 CGTTTTATTCTGCGCCAGAGTCGCGCATAAAAACTGCCAAAATAAAGGGAGAAAAATATG 900
Db    2295 CGTTTTATTCTGCGCCAGAGTCGCGCATAAAAACTGCCAAAATAAAGGGAGAAAAATATG 2354

Qy     901 TTAATTAGTAATGTGGGAATAAATCCCGCCGCTTATTTAAATAATCATTCTGTTGAGAAT 960
Db    2355 TTAATTAGTAATGTGGGAATAAATCCCGCCGCTTATTTAAATAATCATTCTGTTGAGAAT 2414

Qy     961 AGTTCACAGACAGCTTCGCAATCCGTTAGCGCTAAAGATATTCTGAATAGTATTGGTATT 1020
Db    2415 AGTTCACAGACAGCTTCGCAATCCGTTAGCGCTAAAGATATTCTGAATAGTATTGGTATT 2474

Qy    1021 AGCAGCAGTAAAGTCAGTGACCTGGGGTTGAGTCCTACACTGAGCGCGCCTGCGCCAGGG 1080
Db    2475 AGCAGCAGTAAAGTCAGTGACCTGGGGTTGAGTCCTACACTGAGCGCGCCTGCGCCAGGG 2534

Qy    1081 GTATTAACGCAAACCCCCGGAACGATCACGTCCTTTTTAAAAGCCAGTATTCAAAATACC 1140
Db    2535 GTATTAACGCAAACCCCCGGAACGATCACGTCCTTTTTAAAAGCCAGTATTCAAAATACC 2594

Qy    1141 GACATGAATCAGGATTTGAATGCTCTGGCAAATAATGTCACGACTAAAGCGAATGAGGTT 1200
Db    2595 GACATGAATCAGGATTTGAATGCCCTGGCAAATAATGTCACGACTAAAGCGAATGAGGTT 2654

Qy    1201 GTGCAAACCCAGTTACGCGAGCAGCAGGCAGAAGTCGGAAAGTTTTTTGATATTAGCGGA 1260
Db    2655 GTGCAAACCCAGTTACGCGAGCAGCAGGCAGAAGTCGGAAAGTTTTTTGATATTAGCGGA 2714

Qy    1261 ATGTCTTCCAGTGCCGTTGCGCTGTTGGCTGCCGCGAATACGTTAATGCTGACGTTGAAC 1320
Db    2715 ATGTCTTCCAGTGCCGTTGCGCTGTTGGCTGCCGCGAATACGTTAATGCTGACGTTGAAC 2774

Qy    1321 CAGGCTGATAGCAAACTGTCTGGTAAGTTGTCATTAGTCAGTTTTGATGCAGCTAAAACG 1380
Db    2775 CAGGCTGATAGCAAACTGTCTGGTAAGTTGTCATTAGTCAGTTTTGATGCAGCTAAAACG 2834

Qy    1381 ACGGCAAGCTCCATGATGCGCGAAGGGATGAATGCGTTGTCCGGTAGTATTTCCCAGAGC 1440
Db    2835 ACGGCAAGCTCCATGATGCGCGAAGGGATGAATGCGTTGTCCGGTAGTATTTCCCAGAGC 2894

Qy    1441 GCGCTTCAGTTGGGGATCACTGGCGTGGGCGCCAAACTGGAATATAAGGGGCTGCAGAAT 1500
Db    2895 GCGCTTCAGTTGGGGATCACTGGCGTGGGCGCCAAACTGGAATATAAGGGGCTGCAGAAT 2954

Qy    1501 GAAAGAGGCGCGCTTAAACATAATGCCGCGAAGATCGATAAACTGACCACTGAAAGCCAC 1560
Db    2955 GAAAGAGGCGCGCTTAAACATAATGCCGCGAAGATCGATAAACTGACCACTGAAAGCCAC 3014

Qy    1561 AGTATTAAAAAACGTGCTGAACGGGCAGAATAGCGTCAAACTCGGTGCTGAAGGCGTCGAT 1620
Db    3015 AGTATTAAAAAACGTGCTGAACGGGCAGAATAGCGTCAAACTTGGTGCTGAAGGCGTCGAT 3074

Qy    1621 TCTCTGAAATCGTTAAATATGAAGAAAACCGGTACCGATGCGACGAAAAATCTTAATGAT 1680
Db    3075 TCTCTGAAATCGTTAAATATGAAGAAAACCGGTACCGATGCGACGAAAAATCTTAATGAT 3134

Qy    1681 GCGACGCTTAAATCTAATGCCGGAACCAGCGCCACGGAAAGTCTGGGTATTAAAGACAGT 1740
Db    3135 GCGACGCTTAAATCTAATGCCGGAACCAGCGCCACGGAAAGTCTGGGTATTAAAAACAGT 3194

Qy    1741 AATAAACAAATCTCCCCTGAACATCAGGCTATTCTGTCGAAACGTCTTGAGTCTGTCGAA 1800
Db    3195 AATAAACAAATCTCCCCTGAACATCAGGCTATTCTGTCGAAACGTCTTGAGTCTGTCGAA 3254

Qy    1801 TCCGATATTCGTCTTGAGCAGAATACCATGGATATGACCCGAATCGATGCGCGCAAGATG 1860
Db    3255 TCCGATATTCGTCTTGAGCAGAATACCATGGATATGACCCGAATCGATGCGCGCAAGATG 3314

Qy    1861 CAGATGACGGGCGATCTGATTATGAAGAACTCGGTCACGGTCGGTGGTATTGCAGGGGCG 1920
Db    3315 CAGATGACGGGCGATCTGATTATGAAGAACTCAGTCACGGTCGGTGGTATTGCAGGGGCG 3374

Qy    1921 TCCGGGCAGTACGCCGCTACTCAGGAACGTTCCGAGCAGCAAATTAGCCAGGTGAATAAC 1980
Db    3375 TCCAGGCAGTACGCCGCTACTCAGGAACGTTCCGAGCAGCAAATTAGCCAGGTGAATAAC 3434

Qy    1981 CGGGTTGCCAGCACCGCATCGGACGAAGCCCGTGAAAGTTCACGTAAATCGACCAGCCTG 2040
Db    3435 CGGGTTGCCAGCACCGCATCGGACGAAGCCCGTGAAAGTTCACGTAAATCGACCAGCCTG 3494

Qy    2041 ATTCAGGAAATGCTGAAAACAATGGAGAGCATTAACCAGTCGAAAGCATCCGCACTCGCT 2100
Db    3495 ATTCAGGAAATGCTGAAAACAATGGAGAGCATTAACCAGTCGAAAGCATCCGCACTCGCT 3554
```

```
Qy   2101 GCTATCGCAGGCAATATTCGCGCTTAATCTGAAAGGTCATCTATACGCCATCATGGGTGT 2160
          |||||||||||||||||||||||||||||| || ||| ||||||||||||| ||| |
Db   3555 GCTATCGCAGGCAATATTCGCGCTTAATCTGACAGATCAACTATACGCCATCA-GGGGGG 3613

Qy   2161 GATTTAATCGCGCTCCTGATGGCGAACTGGGGATATTATGCTTAATATTCAAAATTATTC 2220
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db   3614 GATTTAATCGCCCTCCTGATGGCGAACTGGGGATATTATGCTTAATATTCAAAATTATTC 3673

Qy   2221 CGCTTCTCCTCATCCGGGGGATCGTTGCCGAACGGCCGCAGACTCCCTCGGCGAGCGAGCA 2280
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db   3674 CGCTTCTCCTCATCCGGGGATCGTTGCCGAACGGCCGCAGACTCCTTCGGCGAGCGAGCA 3733

Qy   2281 CGTCGAGACTGCCGTGGTACCGTCTACCACAGAACATCGCGGTACAGATATCATTTCATT 2340
          || ||||| |||||||||||||||||||||||||||||||||| ||||||||||||||||
Db   3734 CGCCGAGATTGCCGTGGTACCGTCTACCACAGAACATCGCGGCACAGATATCATTTCATT 3793

Qy   2341 ATCGCAGGCGGCTACTAAAATCCACCAGGCACAGCAGACGCTGCAGTCAACGCCACCGAT 2400
          |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db   3794 ATCGCAGGCGGCTACTAAAATCCAGCAGGCACAGCAGACGCTGCAGTCAACGCCACCGAT 3853

Qy   2401 CTCTGAAGAGAATAATGACGAGCGCACGCTGGCGCGCCAGCAGTTGACCAGCAGCCTGAA 2460
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db   3854 TTCTGAAGAGAATAATGACGAGCGCACGCTGGCGCGCCAACAGTTGACCAGCAGCCTGAA 3913

Qy   2461 TGCGCTGGCGAAGTCCGGCGTGTCATTATCCGCAGAACAAAATGAGAACCTGCGGAGCGC 2520
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db   3914 TGCGCTGGCGAAGTCCGGCGTGTCATTATCCGCAGAACAAAATGAGAACCTGCGGAGCAC 3973

Qy   2521 GTTTTCTGCGCCGACGTCGGCCTTATTTAGCGCTTCGCCTATGGCGCAGCCGAGAACAAC 2580
          |||||||||| |||||||||||||||||||||||||||||||  ||   ||||||||||
Db   3974 GTTTTCTGCG-CGACGTCGGCCTTATTTAGCGCTTCGCCTATG--GCCAGCGAGAACAAC 4030

Qy   2581 CATTTCTGATGCTGAGATTTGGGATATGGTTTCCCAAAATATATCGGCGATAGGTGACAG 2640
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4031 CATTTCTGATGCTGAGATTTGGGATATGGTTTCCCAAAATATATCGGCGATAGGTGACAG 4090

Qy   2641 CTATCTGGGCGTTTATGAAAACGTTGTCGCAGTCTATACCGATTTTTATCAGGCCTTCAG 2700
          ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4091 CTACCTGGGCGTTTATGAAAACGTTGTCGCAGTCTATACCGATTTTTATCAGGCCTTCAG 4150

Qy   2701 TGATATTCTTTCCAAAATGGGAGGCTGGTTATTACCAGGTAAGGACGGTAATACCGTTAA 2760
          |||||||||||||||||||||||||||||||  || |||||||| || |||||| ||||
Db   4151 TGATATTCTTTCCAAAATGGGAGGCTGGTTATCGCCTGGTAAGGATGGAAATACCATTAA 4210

Qy   2761 GCTAGATGTTACCTCACTCAAAAATGATTTAAACAGTTTAGTCAATAAATATAATCAAAT 2820
          |||| |||||  |||||| |||| ||| ||| |||||| | |||||||| | ||||||
Db   4211 GCTAAATGTTGACTCACTTAAAAGTGAAATAAGTAGTTTAATTAATAAATACACTCAAAT 4270

Qy   2821 AAACAGTAATACCGTTTTATTTCCAGCGCAGTCAGGCAGCGGCGTTAAAGTAGCCACTGA 2880
          ||| |  ||||| \\\\\\\\\  |||| | |||||||  | | |  ||| ||  |
Db   4271 AAATAAAAATACGATTTTATTTCCCTCGCAAACTGGCAGCGGAATGACAACAGCAACGAA 4330

Qy   2881 AGCGGAAGCGAGACAGTGGCTCAGTGAATTGAATTTACCGAATAGCTGCCTGAAATCTTA 2940
          |||||||||    |||||| | |  |||||||||||||||| | ||||| || ||  | |
Db   4331 AGCGGAAGCTGAGCAGTGGATTAAAGAATTGAATTTACCGGACAGCTGTCTAAAGGCGTC 4390

Qy   2941 TGGATCCGGTTATGTCGTCACCGTTGATCTGACGCCATTACAAAAAATGGTTCAGGATAT 3000
          ||| || |||||||||||    || |||  |  |||| |    |||||||||   ||| |
Db   4391 TGGTTCTGGTTATGTCGTACTGGTGGATACGGGGCCACTGAGCAAAATGGTTAGCGATCT 4450

Qy   3001 TGATGGTTTAGGCGCGCCGGGAAAAGACTCAAAACTCGAAATGGATAACGCCAAATATCA 3060
          | ||||  |||| ||          |  |||   || ||| ||||||||||||||||||
Db   4451 TAATGGAATAGGATCG---------GGTTCAGCCCTTGAACTGGATAACGCCAAATATCA 4501

Qy   3061 AGCCTGGCAGTCGGGTTTTAAAGCGCAGGAAGAAAATATGAAAACCACATTACAGACGCT 3120
          |||||||||||||||||||||| |||||||||||  |||||||||||||||||||||||
Db   4502 AGCCTGGCAGTCGGGTTTTAAAGCACAGGAAGAAAATCTGAAAACCACATTACAGACGCT 4561

Qy   3121 GACGCAAAAATATAGCAATGCCAATTCATTGTACGACAACCTGGTAAAAGTGCTGAGCAG 3180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4562 GACGCAAAAATATAGCAATGCCAATTCATTGTACGACAACCTGGTAAAAGTGCTGAGCAG 4621

Qy   3181 TACGATAAGTAGCAGCCTGGAAACCGCCAAAAGCTTCCTGCAAGGATAACAGAAGAGGAT 3240
          •                                                           •
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4622 TACGATAAGTAGCAGCCTGGAAACCGCCAAAAGCTTCCTGCAAGGATAACAGAAGAGGAT 4681

Qy   3241 ATTAATAATGGTTACAAGTGTAAGGACTCAGCCCCCCGTCATAATGCCAGGTATGCAGAC 3300
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db   4682 ATTAATAATGGTTACAAGTGTAAGAACTCAGCCCCCCGTCATAATGCCAGGTATGCAGAC 4741

Qy   3301 CGAGATCAAAACGCAGGCCACGAATCTTGCGGCGAATCTTTCCGCAGTCAGAGAAAGTGC 3360
          |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db   4742 CGAGATCAAAACGCAGGCCACGAATCTTGCGGCGAATCTTTCCGCGGTCAGAGAAAGTGC 4801

Qy   3361 CACAGCGACGCTGTCAGGGGAAATTAAAGGCCCGCAACTGGAAGATTTTCCCGCGCTGAT 3420
          |||||| ||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   4802 CACAGCAACGCTGTCAGGGGAAATTAAAGGCCAGCAACTGGAAGATTTTCCCGCGCTGAT 4861

Qy   3421 CAAACAGGCGAGTCTGGATGC 3441
          |||||||||||||||||||||
Db   4862 CAAACAGGCGAGTCTGGATGC 4882
```

```
RESULT   3
STU25631
LOCUS       STU25631     3622 bp    DNA             BCT       06-SEP-1995
DEFINITION  Salmonella typhimurium invasin homologs SicA (sicA), SipB (sipB)
            and SipC  (sipC) genes, complete cds.
ACCESSION   U25631
NID         g975293
VERSION     U25631.1  GI:975293
KEYWORDS    .
SOURCE      Salmonella typhimurium.
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Salmonella.
REFERENCE   1  (bases 1 to 3622)
  AUTHORS   Kaniga,K., Tucker,S., Trollinger,D. and Galan,J.E.
  TITLE     Homologs of the Shigella IpaB and IpaC invasins are required for
            Salmonella typhimurium entry into cultured epithelial cells
  JOURNAL   J. Bacteriol. 177 (14), 3965-3971 (1995)
  MEDLINE   95332200
REFERENCE   2  (bases 1 to 3622)
  AUTHORS   Galan,J.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-APR-1995) Jorge E. Galan, Molecular Genetics and
            Microbiology, School of Medicine, SUNY Stony Brook, Stony Brook, NY
            11794-5222, USA
FEATURES             Location/Qualifiers
     source          1. .3622
                     /organism="Salmonella typhimurium"
                     /strain="SL1344"
                     /db_xref="taxon:602"
     gene            75. .572
                     /gene="sicA"
     CDS             75. .572
                     /gene="sicA"
                     /codon_start=1
                     /transl_table=11
                     /product="SicA"
                     /protein_id="AAA75168.1"
                     /db_xref="PID:g975294"
                     /db_xref="GI:975294"
                     /translation="MDYQNNVSEERVAEMIWDAVSEGATLKDVHGIPQDMMDGLYAHA
                     YEFYNQGRLDEAETFFRFLCIYDFYNPDYTMGLAAVCQLKKQFQKACDLYAVAFTLLK
                     NDYRPVFFTGQCQLLMRKAAKARQCFELVNERTEDESLRAKALVYLEALKTAETEQHS
                     EQEKE"
     gene            575. .2356
                     /gene="sipB"
     CDS             575. .2356
                     /gene="sipB"
                     /codon_start=1
                     /transl_table=11
                     /product="SipB"
                     /protein_id="AAA75169.1"
                     /db_xref="PID:g975295"
```

```
                     /db_xref="GI:975295"
                     /translation="MVNDASSISRSGYTQNPRLAEAAFEGVRKNTDFLKAADKAFKDV
                     VATKAGDLKAGTKSGESAINTVGLKPPTDAARERKLSSEGQLTLLLGKLMTLLGDVSLS
                     QLESRLAVWQAMIESQKEMGIQVSKEFQTALGEAQEATDLYEASIKKTDTAKSVYDAA
                     TKKLIQAQNKLQSLDPADPGYAQAEAAVEQAGKEATEAKEALDKATDATVKAGTDAKA
                     KAEKADNILTKFQGTANAASQNQVSQGEQDNLSNVARLTMLMAMFIEIVGKNTEESLQ
                     NDLALFNALQEGRQAEMEKKSAEFQEETRKAEETNRIMGCIGKVLGALLTIVSVVAAV
                     FTGGASLALAAVGLAVMVADEIVKAATGVSFIQQALNPIMEHVLKPLMELIGKAITKA
                     LEGLGVDKKTAEMAGSIVGAIVAAIAMVAVIVVVAVVGKGAAAKLGNALSKMMGETIK
                     KLVPNVLKQLAQNGSKLFTQGMQRITSGLGNVGSKMGLQTNALSKELVGNTLNKVALG
                     MEVTNTAAQSAGGVAEGVFIKNASEALADFMLARFAMDQIQQWLKQSVEIFGENQKVT
                     AELQKAMSSAVQQNADASRFILRQSRA"
     gene            2384. .3613
                     /gene="sipC"
     CDS             2384. .3613
                     /gene="sipC"
                     /codon_start=1
                     /transl_table=11
                     /product="SipC"
                     /protein_id="AAA75170.1"
                     /db_xref="PID:g975296"
                     /db_xref="GI:975296"
                     /translation="MLISNVGINPAAYLNNHSVENSSQTASQSVSAKDILNSIGISSS
                     KVSDLGLSPTLSAPAPGVLTQTPGTITSSLKASIQNTDMNQDLNALANNVTTKANEVV
                     QTQLREQQAEVGKFFDISGMSSSAVALLAAANTLMLTLNQADSKLSGKLSLVSFDAAK
                     TTASSMMREGMNALSGSISQSALQLGITGVGAKLEYKGLQNERGALKHNAAKIDKLTT
                     ESHSIKNVLNGQNSVKLGAEGVDSLKSLNIRKPVPMRRKILMMRRLNLMPEPAPRKVW
                     VLKTVINKVSLNIYILSKRLESVESDIRLEQNYMDITRIDSAQDADDGRSDYEELGHG
                     RWYCRGVRAVRRYSGNVSEQQISQVNNRVASTASDEARESSRKSTSLIQEMLKTMESI
                     NQSKASALAAIAGNIRA"
BASE COUNT     1049 a    759 c   1008 g    806 t
ORIGIN
```

Query Match 60.0%; Score 2063; DB 1; Length 3622;
Best Local Similarity 99.0%; Pred. No. 0;
Matches 2118; Conservative 0; Mismatches 15; Indels 6; Gaps 4;

```
Qy       1 CGCAAAGCCGAGGAAACGAACCGCATTATGGGATGTATCGGGAAAGTCCTCGGCGCGCTG 60
Db    1487 CGCAAAGCCGAGGAAACGAACCGCATTATGGGATGTATCGGGAAAGTCCTCGGCGCGCTG 1546

Qy      61 CTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCT 120
Db    1547 CTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCT 1606

Qy     121 GCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCG 180
Db    1607 GCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCG 1666

Qy     181 TTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTG 240
Db    1667 TTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTG 1726

Qy     241 ATTGGCAAGGCGATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAG 300
Db    1727 ATTGGCAAGGCGATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAG 1786

Qy     301 ATGGCCGGCAGCATTGTTGGTGCGATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTG 360
Db    1787 ATGGCCGGCAGCATTGTTGGTGCGATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTG 1846

Qy     361 GTGGTCGCAGTTGTCGGGAAAGGCGCGGCGGCGAAACTGGGTAACGCGCTGAGCAAAATG 420
Db    1847 GTGGTCGCAGTTGTCGGGAAAGGCGCGGCGGCGAAACTGGGTAACGCGCTGAGCAAAATG 1906

Qy     421 ATGGGCGAAACGATTAAGAAGTTGGTGCCTAACGTGCTGAAACAGTTGGCGCAAAACGGC 480
Db    1907 ATGGGCGAAACGATTAAGAAGTTGGTGCCTAACGTGCTGAAACAGTTGGCGCAAAACGGC 1966

Qy     481 AGCAAACTCTTTACCCAGGGGATGCAACGTATTACTAGCGGTCTGGGTAATGTGGGTAGC 540
Db    1967 AGCAAACTCTTTACCCAGGGGATGCAACGTATTACTAGCGGTCTGGGTAATGTGGGTAGC 2026

Qy     541 AAGATGGGCCTGCAAACGAATGCCTTAAGTAAAGAGCTGGTAGGTAATACCCTAAATAAA 600
Db    2027 AAGATGGGCCTGCAAACGAATGCCTTAAGTAAAGAGCTGGTAGGTAATACCCTAAATAAA 2086

Qy     601 GTGGCGTTGGGCATGGAAGTCACGAATACCGCAGCCCAGTCAGCCGGTGGTGTTGCCGAG 660
Db    2087 GTGGCGTTGGGCATGGAAGTCACGAATACCGCAGCCCAGTCAGCCGGTGGTGTTGCCGAG 2146

Qy     661 GGCGTATTTATTAAAAATGCCAGCGAGGCGCTTGCTGATTTTATGCTCGCCCGTTTTGCC 720
Db    2147 GGCGTATTTATTAAAAATGCCAGCGAGGCGCTTGCTGATTTTATGCTCGCCCGTTTTGCC 2206

Qy     721 ATGGATCAGATTCAGCAGTGGCTTAAACAATCCGTAGAAATATTTGGTGAAAACCAGAAG 780
Db    2207 ATGGATCAGATTCAGCAGTGGCTTAAACAATCCGTAGAAATATTTGGTGAAAACCAGAAG 2266

Qy     781 GTAACGGCGGAACTGCAAAAAGCCATGTCTTCTGCGGTACAGCAAAATGCGGATGCTTCG 840
Db    2267 GTAACGGCGGAACTGCAAAAAGCCATGTCTTCTGCGGTACAGCAAAATGCGGATGCTTCG 2326

Qy     841 CGTTTTATTCTGCGCCAGAGTCGCGCATAAAAACTGCCAAAATAAAGGGAGAAAAATATG 900
Db    2327 CGTTTTATTCTGCGCCAGAGTCGCGCATAAAAACTGCCAAAATAAAGGGAGAAAAATATG 2386

Qy     901 TTAATTAGTAATGTGGGAATAAATCCCGCCGCTTATTTAAATAATCATTCTGTTGAGAAT 960
Db    2387 TTAATTAGTAATGTGGGAATAAATCCCGCCGCTTATTTAAATAATCATTCTGTTGAGAAT 2445

Qy     961 AGTTCACAGACAGCTTCGCAATCCGTTAGCGCTAAAGATATTCTGAATAGTATTGGTATT 1020
Db    2447 AGTTCACAGACAGCTTCGCAATCCGTTAGCGCTAAAGATATTCTGAATAGTATTGGTATT 2506

Qy    1021 AGCAGCAGTAAAGTCAGTGACCTGGGGTTGAGTCCTACACTGAGCGCGCCTGCGCCAGGG 1080
Db    2507 AGCAGCAGTAAAGTCAGTGACCTGGGGTTGAGTCCTACACTGAGCGCGCCTGCGCCAGGG 2566

Qy    1081 GTATTAACGCAAACCCCCGGAACGATCACGTCCTTTTTAAAAGCCAGTATTCAAAATACC 1140
Db    2567 GTATTAACGCAAACCCCCGGAACGATCACGTCCTCTTTAAAAGCCAGTATTCAAAATACC 2626

Qy    1141 GACATGAATCAGGATTTGAATGCTCTGGCAAATAATGTCACGACTAAAGCGAATGAGGTT 1200
Db    2627 GACATGAATCAGGATTTGAATGCTCTGGCAAATAATGTCACGACTAAAGCGAATGAGGTT 2686

Qy    1201 GTGCAAACCCAGTTACGCGAGCAGCAGGCAGAAGTCGGAAAGTTTTTTGATATTAGCGGA 1260
Db    2687 GTGCAAACCCAGTTACGCGAGCAGCAGGCAGAAGTCGGAAAGTTTTTTGATATTAGCGGA 2746

Qy    1261 ATGTCTTCCAGTGCCGTTGCGCTGTTGGCTGCCGCGAATACGTTAATGCTGACGTTGAAC 1320
Db    2747 ATGTCTTCCAGTGCCGTTGCGCTGTTGGCTGCCGCGAATACGTTAATGCTGACGTTGAAC 2806

Qy    1321 CAGGCTGATAGCAAACTGTCTGGTAAGTTGTCATTAGTCAGTTTTGATGCAGCTAAAACG 1380
Db    2807 CAGGCTGATAGCAAACTGTCTGGTAAGTTGTCATTAGTCAGTTTTGATGCAGCTAAAACG 2866

Qy    1381 ACGGCAAGCTCCATGATGCGCGAAGGGATGAATGCGTTGTCCGGTAGTATTTCCCAGAGC 1440
Db    2867 ACGGCAAGCTCCATGATGCGCGAAGGGATGAATGCGTTGTCCGGTAGTATTTCCCAGAGC 2926

Qy    1441 GCGCTTCAGTTGGGGATCACTGGCGTGGGCGCCAAACTGGAATATAAGGGGCTGCAGAAT 1500
Db    2927 GCGCTTCAGTTGGGGATCACTGGCGTGGGCGCCAAACTGGAATATAAGGGGCTGCAGAAT 2986

Qy    1501 GAAAGAGGCGCGCTTAAACATAATGCCGCGAAGATCGATAAACTGACCACTGAAAGCCAC 1560
Db    2987 GAAAGAGGCGCGCTTAAACATAATGCCGCGAAGATCGATAAACTGACCACTGAAAGCCAC 3046

Qy    1561 AGTATTAAAAACGTGCTGAACGGGCAGAATAGCGTCAAACTCGGTGCTGAAGGCGTCGAT 1620
Db    3047 AGTATTAAAAACGTGCTGAACGGGCAGAATAGCGTCAAACTCGGTGCTGAAGGCGTCGAT 3106
```

```
Qy   1621 TCTCTGAAATCGTTAAATATGAAGAAAACCGGTACCGATGCGACGAAAAATCTTAATGAT 1680
          ||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db   3107 TCTCTGAAATCGTTAAATAT-AAGAAAACCGGTACCGATGCGACGAAAAATCTTAATGAT 3165

Qy   1681 GCGACGCTTAAATCTAATGCCGGAACCAGCGCCACGGAAAGTCTGGGTATTAAAGACAGT 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3166 GCGACGCTTAAATCTAATGCCGGAACCAGCGCCACGGAAAGTCTGGGTATTAAAGACAGT 3225

Qy   1741 AATAAACAAATCTCCCCTGAACATCAGGCTATTCTGTCGAAACGTCTTGAGTCTGTCGAA 1800
          |||||||||||    |||||||||||    ||| ||||||||||||||||||||||||||
Db   3226 AATAAACAAAGTCTCCCTGAACATC--TATATCTTGTCGAAACGTCTTGAGTCTGTCGAA 3283

Qy   1801 TCCGATATTCGTCTTGAGCAGAATACCATGGATATGACCCGAATCGAT-GCGCGCAAGAT 1859
          ||||||||||||||||||||||   |||||||| |||||||||||  |||||||||||
Db   3284 TCCGATATTCGTCTTGAGCAGAATTACATGGATATTACCCGAATCGATAGCGCGCAAGAT 3343

Qy   1860 GCAGATGACGGGCGATCTGATTATGAAGAACTCGGTCACGGTCGGTGGTATTGCAGGGGC 1919
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3344 GCAGATGACGGGCGATCTGATTATGAAGAACTCGGTCACGGTCGGTGGTATTGCAGGGGC 3403

Qy   1920 GTCCGGGCAGTACGCCGCTACTCAGG--AACGTTCCGAGCAGCAAATTAGCCAGGTGAAT 1977
          |||||||||||||||||||||||||  |   |||||||||||||||||||||||||||||
Db   3404 GTCCGGGCAGTACGCCGCTACTCAGGTAACGTTTCCGAGCAGCAAATTAGCCAGGTGAAT 3463

Qy   1978 AACCGGGTTGCCAGCACCGCATCGGACGAAGCCCGTGAAAGTTCACGTAAATCGACCAGC 2037
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3464 AACCGGGTTGCCAGCACCGCATCGGACGAAGCCCGTGAAAGTTCACGTAAATCGACCAGC 3523

Qy   2038 CTGATTCAGGAAATGCTGAAAACAATGGAGAGCATTAACCAGTCGAAAGCATCCGCACTC 2097
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3524 CTGATTCAGGAAATGCTGAAAACAATGGAGAGCATTAACCAGTCGAAAGCATCCGCACTC 3583

Qy   2098 GCTGCTATCGCAGGCAATATTCGCGCTTAATCTGAAAGG 2136
          |||||||||||||||||||||||||||||||||||||
Db   3584 GCTGCTATCGCAGGCAATATTCGCGCTTAATCTGAAAGG 3622
```

```
RESULT   4
SDU66877
LOCUS       SDU66877     2995 bp    DNA             BCT       29-AUG-1996
DEFINITION  Salmonella dublin SicA (sicA) and SipB (sipB) genes, complete cds.
ACCESSION   U66877
NID         g1513311
VERSION     U66877.1  GI:1513311
KEYWORDS    .
SOURCE      Salmonella dublin.
  ORGANISM  Salmonella dublin
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Salmonella.
REFERENCE   1  (bases 1 to 2995)
  AUTHORS   Mullan,P.B., Gautier,A.V., Wood,M.W., Edwards,M.H., Jones,B.V. and
            Galyov,E.E.
  TITLE     Sip invasins of Salmonella dublin are involved in macrophage
            cytotoxicity and mouse virulence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2995)
  AUTHORS   Mullan,P.B., Gautier,A.V., Wood,M.W., Edwards,M.H., Jones,B.V. and
            Galyov,E.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-AUG-1996) Institute for Animal Health, Compton,
            Newbury, Berkshire, RG20 7NN, UK
FEATURES             Location/Qualifiers
     source          1. .2995
                     /organism="Salmonella dublin"
                     /strain="2229"
                     /db_xref="taxon:593"
     gene            131. .628
                     /gene="sicA"
     CDS             131. .628
                     /gene="sicA"
                     /codon_start=1
                     /transl_table=11
                     /product="SicA"
                     /protein_id="AAB06794.1"
                     /db_xref="PID:g1513312"
                     /db_xref="GI:1513312"
                     /translation="MDYQNNVSEERVAEMIWDAVSEGATLKDVHGIPQDMMDGLYAHA
                     YEFYNQGRLDEAETFFRFLCIYDFYNPDYTMGLAAVCQLKKQFQKACDLYAVAFTLLK
                     NDYRPVFFTGQCQLLMRKAAKARQCFELVNERTEDESLRAKALVYLEALKTAETEQHS
                     EQEKE"
     gene            631. .2412
                     /gene="sipB"
     CDS             631. .2412
                     /gene="sipB"
                     /function="virulence-associated secreted protein"
                     /codon_start=1
                     /transl_table=11
                     /product="SipB"
                     /protein_id="AAB06795.1"
                     /db_xref="PID:g1513313"
                     /db_xref="GI:1513313"
                     /translation="MVNDASSISRSGYTQNPRLAEAAFEGVRKNTDFLKAADKAFKDV
                     VATKAGDLKAGTKSGESAINTVGLKPPTDAAREKLSSEGQLTLLLGKLMTLLGDVSLS
                     QLESRLAVWQAMIESQKEMGIQVSKEFQTALGEAQEATDLYEASIKKTDTAKSVYDAA
                     TKKLTQAQNKLQSLDPADPGYAQAEAAVEQAGKEATEAKEALDKATDATVKAGTDAKA
                     KAEKADNILTKFQGTANAASQNQVSQGEQDNLSNVARLTMLMAMFIEIVGKNTEESLQ
                     NDLALFNALQEGRQAEMEKKSAEFQEETRKAEETNRIMGCIGKVLGALLTIVSVVAAV
                     FTGGASLALAAVGLAVMVADEIVKAATGVSFIQQALNPIMEHVLKPLMELIGKAITKA
                     LEGLGVDKKTAEMAGSIVGAIVAAIAMVAVIVVVAVVGKGAAAKLGNALSKMMGETIK
                     KLVPNVLKQLAQNGSKLFTQGMQRITSGLGNVGSKMGLQTNALSKELVGNTLNKVALG
                     MEVTNTAAQSAGGVAEGVFIKNASEALADFMLARFAMDQIQQWLKQSVEIFGENQKVT
                     AELQKAMSSAVQQNADASRFILRQSRA"
BASE COUNT      851 a    614 c    847 g    683 t
ORIGIN
```

```
  Query Match           41.9%;  Score 1443.4;  DB 1;  Length 2995;
  Best Local Similarity 99.6%;  Pred. No. 0;
  Matches 1447;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      1 CGCAAAGCCGAGGAAACGAACCGCATTATGGGATGTATCGGGAAAGTCCTCGGCGCGCTG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1543 CGCAAAGCCGAGGAAACGAACCGCATTATGGGATGTATCGGGAAAGTCCTCGGCGCGCTG 1602

Qy     61 CTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1603 CTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCT 1662

Qy    121 GCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1663 GCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGGGTGTCG 1722

Qy    181 TTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1723 TTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTG 1782

Qy    241 ATTGGCAAGGCGATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAG 300
          |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db   1783 ATTGGCAAGGCGATTACCAAAGCGCTGGAAGGGTTAGGCGTCGATAAGAAAACGGCAGAG 1842

Qy    301 ATGGCCGGCAGCATTGTTGGTGCGATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1843 ATGGCCGGCAGCATTGTTGGTGCGATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTG 1902

Qy    361 GTGGTCGCAGTTGTCGGGAAAGGCGCGGCGGCGAAACTGGGTAACGCGCTGAGCAAAATG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1903 GTGGTCGCAGTTGTCGGGAAAGGCGCGGCGGCGAAACTGGGTAACGCGCTGAGCAAAATG 1962

Qy    421 ATGGGCGAAACGATTAAGAAGTTGGTGCCTAACGTGCTGAAACAGTTGGCGCAAAACGGC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1963 ATGGGCGAAACGATTAAGAAGTTGGTGCCTAACGTGCTGAAACAGTTGGCGCAAAACGGC 2022

Qy    481 AGCAAACTCTTTACCCAGGGGATGCAACGTATTACTAGCGGTCTGGGTAATGTGGGTAGC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2023 AGCAAACTCTTTACCCAGGGGATGCAACGTATTACTAGCGGTCTGGGTAATGTAGGTAGC 2082
```

```
Qy    541 AAGATGGGCCTGCAAACGAATGCCTTAAGTAAAGAGCTGGTAGGTAATACCCTAAATAAA 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2083 AAGATGGGCCTGCAAACGAATGCCTTAAGTAAAGAGCTGGTAGGTAATACCCTAAATAAA 2142

Qy    601 GTGGCGTTGGGCATGGAAGTCACGAATACCGCAGCCCAGTCAGCCGGTGGTGTTGCCGAG 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2143 GTGGCGTTGGGCATGGAAGTCACGAATACCGCAGCCCAGTCAGCCGGTGGTGTTGCCGAG 2202

Qy    661 GGCGTATTTATTAAAAATGCCAGCGAGGCGCTTGCTGATTTTATGCTCGCCCGTTTTGCC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2203 GGCGTATTTATTAAAAATGCCAGCGAGGCGCTTGCTGATTTTATGCTCGCCCGTTTTGCC 2262

Qy    721 ATGGATCAGATTCAGCAGTGGCTTAAACAATCCGTAGAAATATTTGGTGAAAACCAGAAG 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2263 ATGGATCAGATTCAGCAGTGGCTTAAACAATCCGTAGAAATATTTGGTGAAAACCAGAAG 2322

Qy    781 GTAACGGCGGAACTGCAAAAAGCCATGTCTTCTGCGGTACAGCAAAATGCGGATGCTTCG 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2323 GTAACGGCGGAACTGCAAAAAGCCATGTCTTCTGCGGTACAGCAAAATGCGGATGCTTCG 2382

Qy    841 CGTTTTATTCTGCGCCAGAGTCGCGCATAAAAACTGCCAAAATAAAGGGAGAAAAATATG 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2383 CGTTTTATTCTGCGCCAGAGTCGCGCATAAAAACTGCCAAAATAAAGGGAGAAAAATATG 2442

Qy    901 TTAATTAGTAATGTGGGAATAAATCCCGCCGCTTATTTAAATAATCATTCTGTTGAGAAT 960
          |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db   2443 TTAATTAGTAATGTGGGAATAAATCCCGCCGCTTATTTAAATAATCATTCAGTTGAGAAT 2502

Qy    961 AGTTCACAGACAGCTTCGCAATCCGTTAGCGCTAAAGATATTCTGAATAGTATTGGTATT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2503 AGTTCACAGACAGCTTCGCAATCCGTTAGCGCTAAAGATATTCTGAATAGTATTGGTATT 2562

Qy   1021 AGCAGCAGTAAAGTCAGTGACCTGGGGTTGAGTCCTACACTGAGCGCGCCTGCGCCAGGG 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2563 AGCAGCAGTAAAGTCAGTGACCTGGGGTTGAGTCCTACACTGAGCGCGCCTGCGCCAGGG 2622

Qy   1081 GTATTAACGCAAACCCCCGGAACGATCACGTCCTTTTTAAAAGCCAGTATTCAAAATACC 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2623 GTATTAACGCAAACCCCCGGAACGATCACGTCCTTTTTAAAAGCCAGTATTCAAAATACC 2682

Qy   1141 GACATGAATCAGGATTTGAATGCTCTGGCAAATAATGTCACGACTAAAGCGAATGAGGTT 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2683 GACATGAATCAGGATTTGAATGCTCTGGCAAATAATGTCACGACTAAAGCGAATGAGGTT 2742

Qy   1201 GTGCAAACCCAGTTACGCGAGCAGCAGGCAGAAGTCGGAAAGTTTTTTGATATTAGCGGA 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2743 GTGCAAACCCAGTTACGCGAGCAGCAGGCAGAAGTCGGAAAGTTTTTTGATATTAGCGGA 2802

Qy   1261 ATGTCTTCCAGTGCCGTTGCGCTGTTGGCTGCCGCGAATACGTTAATGCTGACGTTGAAC 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2803 ATGTCTTCCAGTGCCGTTGCGCTGTTGGCTGCCGCGAATACGTTAATGCTGACGTTGAAC 2862

Qy   1321 CAGGCTGATAGCAAACTGTCTGGTAAGTTGTCATTAGTCAGTTTTGATGCAGCTAAAACG 1380
          |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db   2863 CAGGCTGATAGCAAACTGTCTGGTAAGTTGTCATTAGTCAGCTTTGATGCAGCTAAAACG 2922

Qy   1381 ACGGCAAGCTCCATGATGCGCGAAGGGATGAATGCGTTGTCCGGTAGTATTTCCCAGAGC 1440
          || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2923 ACAGCAAGCTCCATGATGCGCGAAGGGATGAATGCGTTGTCCGGTAGTATTTCCCAGAGC 2982

Qy   1441 GCGCTTCAGITGG 1453
          ||||||||||||
Db   2983 GCGCTTCAGTTGG 2995
```

```
RESULT   5
STU40013
LOCUS       STU40013     3500 bp    DNA             BCT       02-FEB-1996
DEFINITION  Salmonella typhimurium sipD, sipA, and acyl carrier protein (iacP)
            genes, complete cds.
ACCESSION   U40013
NID         g1172126
VERSION     U40013.1  GI:1172126
KEYWORDS    .
SOURCE      Salmonella typhimurium strain=SL1344.
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Salmonella.
REFERENCE   1  (bases 1 to 3500)
  AUTHORS   Kaniga,K., Trollinger,D. and Galan,J.E.
  TITLE     Identification of two targets of the type III protein secretion
            system encoded by the inv and spa loci of Salmonella typhimurium
            that have homology to the Shigella IpaD and IpaA proteins
  JOURNAL   J. Bacteriol. 177 (24), 7078-7085 (1995)
  MEDLINE   96099289
REFERENCE   2  (bases 1 to 3500)
  AUTHORS   Galan,J.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-NOV-1995) Jorge E. Galan, Molecular Gentics and
            Microbiology, School of Medicine, SUNY Stony Brook, Stony Brook, NY
            11794-5222, USA
FEATURES             Location/Qualifiers
     source          1. .3500
                     /organism="Salmonella typhimurium"
                     /strain="SL1344"
                     /db_xref="taxon:602"
     gene            84. .1130
                     /gene="sipD"
     CDS             84. .1130
                     /gene="sipD"
                     /note="homolog of Shigella spp. IpaD"
                     /codon_start=1
                     /transl_table=11
                     /product="SipD"
                     /protein_id="AAA86617.1"
                     /db_xref="PID:g1172127"
                     /db_xref="GI:1172127"
                     /translation="MLNIQNYSASPHPGIVAERPQTPSASEHVETAVVPSTTEHRGTD
                     IISLSQAATKIHQAQQTLQSTPPISEENNDERTLARQQLTSSLNALAKSGVSLSAEQN
                     ENLRSAFSAPTSALFSASPMAQPRTTISDAEIWDMVSQNISAIGDSYLGVYENVVAVY
                     TDFYQAFSDILSKMGGWLLPGKDGNTVKLDVTSLKNDLNSLVNKYNQINSNTVLFPAQ
                     SGSGVKVATEAEARQWLSELNLPNSCLKSYGSGYVVTVDLTPLQKMVQDIDGLGAPGK
                     DSKLEMDNAKYQAWQAGFKAQEENMKTTLQTLTQKYSNANSLYDNLVKVLSSTISSSW
                     KPPKASCKDNRRGY"
     gene            1133. .3187
                     /gene="sipA"
     CDS             1133. .3187
                     /gene="sipA"
                     /note="homolog of Shigella spp. IpaA"
                     /codon_start=1
                     /transl_table=11
                     /product="SipA"
                     /protein_id="AAA86618.1"
                     /db_xref="PID:g1172128"
                     /db_xref="GI:1172128"
                     /translation="MVTSVRTQPPVIMPGMQTEIKTQATNLAANLSAVRAKCHSDAVR
                     EIKGPQLEDFPALIKQASLDALFKCGKDAEALKEVFTNSNNVAGKKAIMEFAGLFRSA
                     LNATSDSPEAKTLLMKVGAEYAAQIIKDGLKEKSAFGPWLPETKKAEAKLENLEKQLL
                     DIIKNNTGGELSKLSTNLVMQEVMPYIASCIEHNFGCTLDPLTRSNLIHLVDKAAAKA
                     VEALDMCPQKLTQEQGTSVGREARHLEMQTLIPLLLRNVFAQIPADKLPDPKIPEPAA
                     GPVPDGGKKAEPTGINININIDSSNHSVDNSKHINNAEPVDNGQRHIDNSNHDNSRKT
                     IDNSRTFIDNSQRNGESHHSTNSSNVSHSHSRVDSTTHQTETAHSASTGAIDHGIAGK
                     IDVTAHATAEAVTNASSESKDGKVVTSEKGTTGETTSFDEVDGVTSKSIIGKPVQATV
                     HGVDDNKQQSQTAEIVNVKPLASQLAGVENVKTDTLQSDTTVITGNKAGTTDNDNSQT
                     DKTGPFSGLKFKQNSFLSTVPSVTNMHSMHFDARETFLGVIRKALEPDTSTPFPVRRA
                     FDGLRAEILPNDTIKSAALKAQCSDIDKHPELKAKMETLKEVITHHPQKEKLAEIALQ
                     FAREAGLTRLKGETDYVLSNVLDGLIGDGSWRAGPAYESYLNKPGVDRVITTVDGLHM
                     QR"
     gene            3206. .3454
                     /gene="iacP"
     CDS             3206. .3454
                     /gene="iacP"
                     /note="IacP"
                     /codon_start=1
                     /transl_table=11
                     /product="acyl carrier protein"
```

```
                     /protein_id="AAA86619.1"
                     /db_xref="PID:g1172129"
                     /db_xref="GI:1172129"
                     /translation="MNMDIEARVKKVITSCIAVDVDSINGQTHLVEDLYADSLDLIDI
                     VFGLSEEFDISCNENDLPDMMTFADICRVVKKSLESRV"
BASE COUNT     1048 a    759 c    879 g    814 t
ORIGIN


  Query Match             37.5%;  Score 1290.4;  DB 1;  Length 3500;
  Best Local Similarity   99.7%;  Pred. No. 0;
  Matches 1324;  Conservative    0;  Mismatches    1;  Indels    3;  Gaps    3;

Qy   2115 TATTCGCGCTTAATCTGAAAGGTCATCTATACGCCATCATGGGTGTGATTTAATCGCGCT 2174
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TATTCGCGCTTAATCTGAAAGGTCATCTATACGCCATCATGGGTGTGATTTAATCGCGCT 60

Qy   2175 CCTGATGGCGAACTGGGGATATTATGCTTAATATTCAAAATTATTCCGCTTCTCCTCATC 2234
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CCTGATGGCGAACTGGGGATATTATGCTTAATATTCAAAATTATTCCGCTTCTCCTCATC 120

Qy   2235 CGGGGATCGTTGCCGAACGGCCGCAGACTCCCTCGGCGAGCGAGCACGTCGAGACTGCCG 2294
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CGGGGATCGTTGCCGAACGGCCGCAGACTCCCTCGGCGAGCGAGCACGTCGAGACTGCCG 180

Qy   2295 TGGTACCGTCTACCACAGAACATCGCGGTACAGATATCATTTCATTATCGCAGGCGGCTA 2354
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TGGTACCGTCTACCACAGAACATCGCGGTACAGATATCATTTCATTATCGCAGGCGGCTA 240

Qy   2355 CTAAAATCCACCAGGCACAGCAGACGCTGCAGTCAACGCCACCGATCTCTGAAGAGAATA 2414
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CTAAAATCCACCAGGCACAGCAGACGCTGCAGTCAACGCCACCGATCTCTGAAGAGAATA 300

Qy   2415 ATGACGAGCGCACGCTGGCGCGCCAGCAGTTGACCAGCAGCCTGAATGCGCTGGCGAAGT 2474
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 ATGACGAGCGCACGCTGGCGCGCCAGCAGTTGACCAGCAGCCTGAATGCGCTGGCGAAGT 360

Qy   2475 CCGGCGTGTCATTATCCGCAGAACAAAATGAGAACCTGCGGAGCGCGTTTTCTGCGCCGA 2534
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CCGGCGTGTCATTATCCGCAGAACAAAATGAGAACCTGCGGAGCGCGTTTTCTGCGCCGA 420

Qy   2535 CGTCGGCCTTATTTAGCGCTTCGCCTATGGCGCAGCCGAGAACAACCATTTCTGATGCTG 2594
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CGTCGGCCTTATTTAGCGCTTCGCCTATGGCGCAGCCGAGAACAACCATTTCTGATGCTG 480

Qy   2595 AGATTTGGGATATGGTTTCCCAAAATATATCGGCGATAGGTGACAGCTATCTGGGCGTTT 2654
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 AGATTTGGGATATGGTTTCCCAAAATATATCGGCGATAGGTGACAGCTATCTGGGCGTTT 540

Qy   2655 ATGAAAACGTTGTCGCAGTCTATACCGATTTTTATCAGGCCTTCAGTGATATTCTTTCCA 2714
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 ATGAAAACGTTGTCGCAGTCTATACCGATTTTTATCAGGCCTTCAGTGATATTCTTTCCA 600

Qy   2715 AAATGGGAGGCTGGTTATTACCAGGTAAGGACGGTAATACCGTTAAGCTAGATGTTACCT 2774
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 AAATGGGAGGCTGGTTATTACCAGGTAAGGACGGTAATACCGTTAAGCTAGATGTTACCT 660

Qy   2775 CACTCAAAAATGATTTAAACAGTTTAGTCAATAAATATAATCAAATAAACAGTAATACCG 2834
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 CACTCAAAAATGATTTAAACAGTTTAGTCAATAAATATAATCAAATAAACAGTAATACCG 720

Qy   2835 TTTTATTTCCAGCGCAGTCAGGCAGCGGCGTTAAAGTAGCCACTGAAGCGGAAGCGAGAC 2894
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 TTTTATTTCCAGCGCAGTCAGGCAGCGGCGTTAAAGTAGCCACTGAAGCGGAAGCGAGAC 780

Qy   2895 AGTGGCTCAGTGAATTGAATTTACCGAATAGCTGCCTGAAATCTTATGGATCCGGTTATG 2954
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 AGTGGCTCAGTGAATTGAATTTACCGAATAGCTGCCTGAAATCTTATGGATCCGGTTATG 840

Qy   2955 TCGTCACCGTTGATCTGACGCCATTACAAAAAATGGTTCAGGATATTGATGGTTTAGGCG 3014
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 TCGTCACCGTTGATCTGACGCCATTACAAAAAATGGTTCAGGATATTGATGGTTTAGGCG 900

Qy   3015 CGCCGGGAAAAGACTCAAAACTCGAAATGGATAACGCCAAATATCAAGCCTGGCAGTCGG 3074
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db    901 CGCCGGGAAAAGACTCAAAACTCGAAATGGATAACGCCAAATATCAAGCCTGGCAGGCGG 960

Qy   3075 GTTTTAAAGCGCAGGAAGAAAATATGAAAACCACATTACAGACGCTGACGCAAAAATATA 3134
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 GTTTTAAAGCGCAGGAAGAAAATATGAAAACCACATTACAGACGCTGACGCAAAAATATA 1020

Qy   3135 GCAATGCCAATTCATTGTACGACAACCTGGTAAAAGTGCTGAGCAGTACGATAAGTAGCA 3194
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 GCAATGCCAATTCATTGTACGACAACCTGGTAAAAGTGCTGAGCAGTACGATAAGTAGCA 1080

Qy   3195 GCCTGGAAACCGCCAAAAGCTTCCTGCAAGGATAACAGAAGAGGATATTAATAATGGTTA 3254
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 G-CTGGAAACCGCCAAAAGCTTCCTGCAAGGATAACAGAAGAGGATATTAATAATGGTTA 1139

Qy   3255 CAAGTGTAAGGACTCAGCCCCCCGTCATAATGCCAGGTATGCAGACCGAGATCAAAACGC 3314
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1140 CAAGTGTAAGGACTCAGCCCCCCGTCATAATGCCAGGTATGCAGACCGAGATCAAAACGC 1199

Qy   3315 AGGCCACGAATCTTGCGGCGAATCTTTCCGCAGTCAGAG-AAAGTGCCACAGCGACGCTG 3373
           |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db   1200 AGGCCACGAATCTTGCGGCGAATCTTTCCGCAGTCAGAGCAAAGTGCCACAGCGACGCTG 1259

Qy   3374 TCAGGGGAAATTAAAGGCCCGCAACTGGAAGATTTTCCCGCGCTGATCAAACAGGCGAGT 3433
           ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1260 TCA-GGGAAATTAAAGGCCCGCAACTGGAAGATTTTCCCGCGCTGATCAAACAGGCGAGT 1318

Qy   3434 CTGGATGC 3441
           ||||||||
Db   1319 CTGGATGC 1326


RESULT   6
SDIPABCD
LOCUS       SDIPABCD     6869 bp    DNA             BCT       24-JUL-1991
DEFINITION  Shigella dysenteriae ipaBCD genes.
ACCESSION   X60777
NID         g46932
VERSION     X60777.1  GI:46932
KEYWORDS    ipaB gene; ipaC gene; ipaD gene; ippI gene.
SOURCE      Shigella dysenteriae.
  ORGANISM  Shigella dysenteriae
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Shigella.
REFERENCE   1  (bases 1 to 6869)
  AUTHORS   Palchaudhuri,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUN-1991) S. Palchaudhuri, Wayne State Univers.School
            of Medicine, Department of Immunology and, Microbiology, 540 E.
            Canfield, Detroit, MI 48201, USA
REFERENCE   2  (bases 1 to 6869)
  AUTHORS   Yao,R. and Palchaudhuri,S.
  TITLE     Nucleotide sequence of the ipaBCD structure genes of Shigella
            dysenteriae
  JOURNAL   Unpublished
FEATURES             Location/Qualifiers
     source          1. .6869
                     /organism="Shigella dysenteriae"
                     /strain="CG097"
                     /db_xref="taxon:622"
     CDS             317. .943
                     /note="24kD ORF"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA43188.1"
                     /db_xref="PID:g46933"
                     /db_xref="GI:46933"
                     /db_xref="SWISS-PROT:Q04314"
                     /translation="MQILNKILPQVEFAIPRPSFNSLSRNKLVKKILSVFNLKQRFPQ
                     KNFGCTVNINKIRDNVIDKIKDSNSGNQLFCWMSQERTTYVSSMINRSIDEMAIHNGV
                     VLTSDNKKNIFAAIEKKFPDIKLDEKSAQTSISHTALNEIASSGLRAKILKRYSSDMD
                     LFNTQMKDLTNLVSSSVYDKIFNESTKVLQIEISAEVLKAVYRQSNTN"
```

```
     CDS             1000. .1467
                     /note="IppI ORF"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA43189.1"
                     /db_xref="PID:g46934"
                     /db_xref="GI:46934"
                     /db_xref="SWISS-PROT:P18008"
                     /translation="MSLNITENESISTAVIDAINSGATLKDINAIPDDMMDDIYSYAY
                     DFYNKGRIEEAEVFFRFLCIYDFYNVDYIMGLAAIYQIKEQFQQAADLYAVAFALGKN
                     DYTPVFHTGQCQLRLKAPLKAKECFELVIQHSNDEKLKIKAQSYLDAIQDIKE"
     CDS             1473. .3215
                     /note="IpaB ORF"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA43190.1"
                     /db_xref="PID:g46935"
                     /db_xref="GI:46935"
                     /db_xref="SWISS-PROT:Q03945"
                     /translation="MHNVNTTTGLSLAKILASTELGDNTIQAGNDAANKLFSLTIAD
                     LTANKNINTTNAHSTSNILIPELKAPKSLNASSQLTLLIGNLIQILGEKSLTALTNKI
                     TAWKSQQQARQQKNLEFSDKINTLLSETEGLTRDYEKQINKLKNADSKIKDLENKINQ
                     IQTRLSELDPDSPEKKKLSREEIQLTIKKDAAVKDRTLIEQKTLSIHSKLTDKSMQLE
                     KEIDSFSAFSNTASAEQLSTQQKSLTGLASVTQLMATFIQLVGKNNEESLKNDLALFQ
                     SLQESRKTEMERKSDEYAAEVRKAEELNRVMGCVGKILGALLTIVSVVAAAFSGGASL
                     ALADVGLALMVTDAIVQAATGNSFMEQALNPIMKAVIEPLIKLLSDAFTKMLEGLGVD
                     SKKAKMIGSILGAIAGALVLVAAVVLVATVGKQAAAKLAENIGKIIGKTLTDLIPKFL
                     KNFSSQLDDLITNAVARLNKFLGAAGDEVISKQIISTHLNQAVLLGESVNSATQAGGS
                     VASAVFQNSASTNLADLTLSKYQVEQLSKYISEAIEKFGQLQEVIADLLASMSNSQAN
                     RTDVAKAILQQTTA"
     CDS             3178. .4326
                     /note="IpaC ORF"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA43191.1"
                     /db_xref="PID:g46936"
                     /db_xref="GI:46936"
                     /db_xref="SWISS-PROT:Q03946"
                     /translation="MLQKQFCNKLLLDTNKENVMEIQNTKSTQILYTDISTKQTQSSS
                     ETQKSQNYQQIAAHIPLNVGKNPVLTTTLNDDQLLKLSEQVQHDSEIIARLTDKKMKD
                     LSEMSHTLTPENTLDISSLSSNAVSLIISVAVLLSALRTAETKLGSQLSLIAFDATKS
                     AAENIVRQGLAALSSSITGAVTQVGITGIGAKKTHSGISDQKGALRKNLATAQSLEKE
                     LAGSKLGLNKQIDTNITSPQTNSSTKFLGKNKLAPDNISLSTEHKTSLSSPDISLQDK
                     IDTQRRTYELNTLSAQQKQNIGRATMETSAVAGNISTSGGRYASALEEEEQLISQASS
                     KQAEEASQVSKEASQATNQLIQKLLNIIDNINQSRSSTASQIAGNIRA"
     CDS             4377. .5375
                     /note="IpaD ORF"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA43192.1"
                     /db_xref="PID:g46937"
                     /db_xref="GI:46937"
                     /db_xref="SWISS-PROT:Q03947"
                     /translation="MNITTLTNSISTSSFSPNNTNGSSTETVNSDIKTTTSSHPVSSL
                     TMLNDTLHNIRTTNQALKKELSQKTLTKTSLEEIALHSSQISMDVNKSAQLLNILSKT
                     EYPINKDARELLHSAPKEAELDGYEMISHRELWAKIANSINDINEQYLKVYEHAVSSY
                     TQMYQEFSAVLSSLAGWISPGGNDGNSVKLQVKSLKDALTTLKKNYEDKPLYPATNTV
                     SEQEANKWLTELGGTIGTVSAKNGGYVVSINMTPIYNMLNRLDNLGGNGEVVLDNAKY
                     QAWNAGFSAEDETMKNNLQTLVQKYSNANSIFDNLVKVLSSTISSCTDTDKLFLHF"
BASE COUNT     2518 a   1369 c   1135 g   1847 t
ORIGIN


  Query Match             5.2%;  Score 179;  DB 1;  Length 6869;
  Best Local Similarity   51.2%;  Pred. No. 4.3e-35;
  Matches  459;  Conservative    0;  Mismatches  420;  Indels   18;  Gaps    1;

Qy       1 CGCAAAGCCGAGGAAACGAACCGCATTATGGGATGTATCGGGAAAGTCCTCGGCGCGCTG 60
           || ||||| || |||   ||| |  | ||||| ||| | |||||| | || || || ||
Db    2364 CGTAAAGCAGAAGAACTCAACAGAGTAATGGGTTGTGTTGGGAAAATACTTGGGGCACTT 2423

Qy      61 CTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCT 120
           |||| || || || ||||| || || | |||| | || || ||   ||| || |||||

Db    2424 TTAACTATCGTTAGTGTTGTTGCAGCAGCTTTTTCTGGAGGAGCCTCTCTAGCACTGGCA 2483

Qy     121 GCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCG 180
           |  || ||  | ||  | |||||  | ||||  || ||  | || ||||| ||    ||
Db    2484 GATGTTGGTTTAGCTCTTATGGTTACGGATGCTATAGTACAAGCAGCGACCGGCAATTCC 2543

Qy     181 TTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTG 240
           || ||  | || || || || ||||| ||| |    ||  |  | || ||||| | ||
Db    2544 TTCATGGAACAAGCCCTGAATCCGATCATGAAAGCAGTCATTGAACCCTTAATCAAACTC 2603

Qy     241 ATTGGCAAGGCGATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAG 300
           ||    | ||  |||| ||| ||| ||||| || |||||||||   |||||  ||  |
Db    2604 CTTTCAGATGCATTTACAAAAATGCTCGAAGGCTTGGGCGTCGACTCGAAAAAAGCCAAA 2663

Qy     301 ATGGCCGGCAGCATTGTTGGTGCGATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTG 360
           |||   |||   ||| | || || || |  | |||| ||| | |||| || |    | ||
Db    2664 ATGATTGGCTCTATTCTGGGGGCAATCGCAGGCGCTCTTGTCCTGGTTGCAGCAGTCGTT 2723

Qy     361 GTGGTCGCAGTTGTCGGGAAAGGCGCGGCGGCGAAACTGGGTAACGCGCTGAGCAAAATG 420
           | || ||   ||| || |||   || || || ||||| |   |     |  |||||||
Db    2724 CTCGTAGCCACTGTTGGTAAACAGGCAGCAGCAAAACTTGCAGAGAATATTGGCAAAATA 2783

Qy     421 ATGGGCGAAACGATTAAGAAGTTGGTGCCTAACGTGCTGAAACAGTTGGCGCAAAACGGC 480
           || ||  |||| | |   |  |  | || ||  | || ||  | ||  |     |
Db    2784 ATAGGTAAAACCCTCACAGACCTTATACCAAAGTTTCTCAAGAATTTTTCTTCTCAACTG 2843

Qy     481 AGCAAACTCTTTACCCAGGGGATGCAACGTATTACTAGCGGTCTGGGTAATGTGGGTAGC 540
           | |  |  | ||  | |   |     |  | | ||    ||| |||   | ||||
Db    2844 GACGATTTAATCACTAATGCTGTTGCCAGATTAAATAAATTTCTTGGTGCAGCGGGTGAT 2903

Qy     541 AAGATGGGCCTGCAAACGAATGCCTTAAGTAAAGAGCTGGTAGGTAATACCCTAAATAAA 600
           |  |                    ||  ||| |  |  |    |     |||  ||
Db    2904 GAAGT-----------------AATATCCAAACAAATTATTTCCACCCATTTAAACCAA 2945

Qy     601 GTGGCGTTGGGCATGGAAGTCACGAATACCGCAGCCCAGTCAGCCGGTGGTGTTGCCGAG 660
           |  |  ||       |||      ||  | ||  | ||  |  |  ||  |||| ||
Db    2946 GCAGTTTTATTAGGAGAAAGTGTTAACTCTGCCACACAAGCGGGAGGAAGTGTCGCTTCT 3005

Qy     661 GGCGTATTTATTAAAAATGCCAGCGAGGCGCTTGCTGATTTTATGCTCGCCCGTTTTGCC 720
           |  || ||    || |  ||          || || ||  | |   |  |    | |
Db    3006 GCTGTTTTCCAGAACAGCGCGTCGACAAATCTAGCAGACCTGACATTATCGAAATATCAA 3065

Qy     721 ATGGATCAGATTCAGCAGTGGCTTAAACAATCCGTAGAAATATTTGGTGAAAACCAGAAG 780
           | || ||  |    | |  | |  || | |||||| ||| ||  ||   ||| |
Db    3066 GTTGAACAACTGTCAAAATATATCAGTGAAGCAATAGAAAAATTCGGCCAATTGCAGGAA 3125

Qy     781 GTAACGGCGGAACTGCAAAAAGCCATGTCTTCTGCGGTACAGCAAAATGCGGATGCTTCG 840
           ||||  || || ||    |  | |||||     |     | |   | |||| | |
Db    3126 GTAATTGCAGATCTATTAGCCTCAATGTCCAACTCTCAGGCTAATAGAACTGATGTTGCA 3185

Qy     841 CGTTTTATTCTGCGCCAGAGTCGCGCATAAAAACTGCCAAAATAAAGGGAGAAAAAT 897
           |||  |||  || | |   || | | |       |  | || |  |   ||||
Db    3186 AAAGCAATTTTGCAACAAACTACTGCTTGATACAAATAAGGAGAATGTTATGGAAAT 3242


RESULT   7
SFPVR2
LOCUS       SFPVR2       7663 bp    DNA             BCT       12-SEP-1993
DEFINITION  Shigella flexineri plasmid YSH6000 virulence region-2 DNA
            containing 24k gene, 18k gene, ipaB gene, ipaC gene and ipaD gene.
ACCESSION   X15319
NID         g47052
VERSION     X15319.1  GI:47052
KEYWORDS    ipaB gene; ipaC gene; ipaD gene; virulence proteins.
SOURCE      Enterobacteriaceae.
  ORGANISM  Enterobacteriaceae
            Bacteria; Proteobacteria; gamma subdivision.
REFERENCE   1  (bases 1 to 7663)
  AUTHORS   Sasakawa,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAY-1989) Sasakawa C., Institute of Medical Science,
            University of Tokyo, Shirokanedai 4-6-1, Minato-ku Tokyo 108, Japan
REFERENCE   2  (bases 1010 to 6538)
```

```
  AUTHORS   Sasakawa,C., Adler,B., Tobe,T., Okada,N., Nagai,S., Komatsu,K. and
            Yoshikawa,M.
  TITLE     Functional organization and nucleotide sequence of virulence
            Region-2 on the large virulence plasmid in Shigella flexneri 2a
  JOURNAL   Mol. Microbiol. 3 (9), 1191-1201 (1989)
  MEDLINE   90014179
COMMENT     Region 2 has three promoters that transcribe RNAs of 4.0, 4.5 and
            7.5kb. The 4.0 kb transcript produces the 24,18kD, IpaB and IpaC
            proteins. The 4.5 kb transcript produces the ipaD protein. The
            7.5kb transcript covers the entire region 2.
            Data kindly reviewed (20-NOV-1989) by Sasakawa C.
FEATURES             Location/Qualifiers
     source          1. .7663
                     /organism="Enterobacteriaceae"
                     /plasmid="YSH6000"
                     /db_xref="taxon:543"
     promoter        1291. .1296
                     /note="-35 region for 4kb transcript"
     promoter        1312. .1318
                     /note="-10 region for 4kb transcript"
     RBS             1360. .1362
                     /note="put. ribosome binding site"
     CDS             1370. .1996
                     /note="24k gene (AA 1-208)"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA33379.1"
                     /db_xref="PID:g47053"
                     /db_xref="GI:47053"
                     /db_xref="SWISS-PROT:P33548"
                     /translation="MQILNKILPQVEFAIPRPSFDSLSRNKLVKKILSVFNLKQRFPQ
                     KNFGCPVNINKIRDSVIDKIKDSNSGNQLFCWMSQERTTYVSSMINRSIDEMAIHNGV
                     VLTSDNKKNIFAAIEKKFPDIKLDEKSAQTSISHTALNEIASSGLRAKILKRYSSDMD
                     LFNTQMKDLTNLVSSSVYDKIFNESTKVLQIEISAEVLKAVYRQSNTN"
     RBS             2043. .2045
                     /note="put. ribosome binding site"
     CDS             2053. .2520
                     /note="18k gene (AA 1-155)"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA33380.1"
                     /db_xref="PID:g47054"
                     /db_xref="GI:47054"
                     /db_xref="SWISS-PROT:P18008"
                     /translation="MSLNITENESISTAVIDAINSGATLKDINAIPDDMMDDIYSYAY
                     DFYNKGRIEEAEVFFRFLCIYDFYNVDYIMGLAAIYQIKEQFQQAADLYAVAFALGKN
                     DYTPVFHTGQCQLRLKAPLKAKECFELVIQHSNDEKLKIKAQSYLDAIQDIKE"
     RBS             2513. .2515
                     /note="put. ribosome binding site"
     CDS             2526. .4268
                     /note="ipaB gene (AA 1-580)"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA33381.1"
                     /db_xref="PID:g47055"
                     /db_xref="GI:47055"
                     /db_xref="SWISS-PROT:P18011"
                     /translation="MHNVSTTTTGFPLAKILASTELGDNTIQAANDAANKLFSLTIAD
                     LTANQNINTTNAHSTSNILIPELKAPKSLNASSQLTLLIGNLIQILGEKSLTALTNKI
                     TAWKSQQQARQQKNLEFSDKINTLLSETEGLTRDYEKQINKLKNADSKIKDLENKINQ
                     IQTRLSELDPESPEKKKLSREEIQLTIKKDAAVKDRTLIEQKTLSIHSKLTDKSMQLE
                     KEIDSFSAFSNTASAEQLSTQQKSLTGLASVTQLMATFIQLVGKNNEESLKNDLALFQ
                     SLQESRKTEMERKSDEYAAEVRKAEELNRVMGCVGKILGALLTIVSVVAAAFSGGASL
                     ALAAVGLALMVTDAIVQAATGNSFMEQALNPIMKAVIEPLIKLLSDAFTKMLEGLGVD
                     SKKAKMIGSILGAIAGALVLVAAVVLVATVGKQAAAKLAENIGKIIGKTLTDLIPKFL
                     KNFSSQLDDLITNAVARLNKFLGAAGDEVISKQIISTHLNQAVLLGESVNSATQAGGS
                     VASAVFQNSASTNLADLTLSKYQVEQLSKYISEAIEKFGQLQEVIADLLASMSNSQAN
                     RTDVAKAILQQTTA"
     RBS             4224. .4226
                     /note="put. ribosome binding site"
     CDS             4231. .5379
                     /note="ipaC gene (AA 1-382)"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA33382.1"
                     /db_xref="PID:g47056"
                     /db_xref="GI:47056"
                     /db_xref="SWISS-PROT:P18012"
                     /translation="MLQKQFCNKLLLDTNKENVMEIQNTKPTQILYTDISTKQTQSSS
                     ETQKSQNYQQIAAHIPLNVGKNPVLTTTLNDDQLLKLSEQVQHDSEIIARLTDKKMKD
                     LSEMSHTLTPENTLDISSLSSNAVSLIISVAVLLSALRTAETKLGSQLSLIAFDATKS
                     AAENIVRQGLAALSSSITGAVTQVGITGIGAKKTHSGISDQKGALRKNLATAQSLEKE
                     LAGSKLGLNKQIDTNITSPQTNSSTKFLGKNKLAPDNISLSTEHKTSLSSPDISLQDK
                     IDTQRRTYELNTLSAQQKQNIGRATMETSAVAGNISTSGGRYASALEEEEQLISQASS
                     KQAEEASQVSKEASQATNQLIQKLLNIIDSINQSKNSTASQIAGNIRA"
     promoter        5376. .5381
                     /note="-35 region for 4.5kb transcript"
     promoter        5396. .5401
                     /note="-10 region for 4.5kb transcript"
     RBS             5421. .5423
                     /note="put. ribosome binding site"
     CDS             5430. .6428
                     /note="ipaD gene (AA 1-332)"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA33383.1"
                     /db_xref="PID:g47057"
                     /db_xref="GI:47057"
                     /db_xref="SWISS-PROT:P18013"
                     /translation="MNITTLTNSISTSSFSPNNTNGSSTETVNSDIKTTTSSHPVSSL
                     TMLNDTLHNIRTTNQALKKELSQKTLIKTSLEEIALHSSQISMDVNKSAQLLDILSRH
                     EYPINKDARELLHSAPKEAELDGDQMISHRELWAKIANSINDINEQYLKVYEHAVSSY
                     TQMYQDFSAVLSSLAGWISPGGNDGNSVKLQVNSLKKALEELKEKYKDKPLYPANNTV
                     SQEQANKWLTELGGTIGKVSQKNGGYVVSINMTPIDNMLKSLDNLGGNGEVVLDNAKY
                     QAWNAGFSAEDETMKNNLQTLVQKYSNANSIFDNLVKVLSSTISSCTDTDKLFLHF"
BASE COUNT     2845 a   1501 c   1246 g   2071 t
ORIGIN
```

Query Match 5.2%; Score 179; DB 1; Length 7663;
Best Local Similarity 51.2%; Pred. No. 4.4e-35;
Matches 459; Conservative 0; Mismatches 420; Indels 18; Gaps 1;

```
Qy       1 CGCAAAGCCGAGGAAACGAACCGCATTATGGGATGTATCGGGAAAGTCCTCGGCGCGCTG 60
           || ||||| || |||   ||| |  | |||||  ||| | ||||||  | || || || ||
Db    3417 CGTAAAGCAGAAGAACTCAACAGAGTAATGGGTTGTGTTGGGAAAATACTTGGGGCACTT 3476

Qy      61 CTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCT 120
            |||| || || || ||||| || || | |||| | || || ||   ||| || |||||
Db    3477 TTAACTATCGTTAGTGTTGTTGCAGCAGCTTTTTCTGGAGGAGCCTCTCTAGCACTGGCA 3536

Qy     121 GCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCG 180
           || || ||  | ||  | |||||  | |||  || ||  | || |||||  ||    ||
Db    3537 GCTGTTGGTTTAGCTCTTATGGTTACGGATGCTATAGTACAAGCAGCGACCGGCAATTCC 3596

Qy     181 TTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTG 240
           || ||  | || || || || ||||| ||| |    ||  |  | || |||||  | ||
Db    3597 TTCATGGAACAAGCCCTGAATCCGATCATGAAAGCAGTCATTGAACCCTTAATCAAACTC 3656

Qy     241 ATTGGCAAGGCGATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAG 300
           ||    | ||  |||| |||  ||| |||||  || |||||||    |||||  ||  |
Db    3657 CTTTCAGATGCATTTACAAAAATGCTCGAAGGCTTGGGCGTCGACTCGAAAAAAGCCAAA 3716

Qy     301 ATGGCCGGCAGCATTGTTGGTGCGATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTG 360
           |||   |||   ||| | || || || |  | |||| ||| | | || || |   | ||
Db    3717 ATGATTGGCTCTATTCTGGGGGCAATCGCAGGCGCTCTTGTCCTAGTTGCAGCAGTCGTT 3776

Qy     361 GTGGTCGCAGTTGTCGGGAAAGGCGCGGCGGCGAAACTGGGTAACGCGCTGAGCAAAATG 420
           | || ||   ||| || |||   || || || ||||| |   |   |  |||||||
Db    3777 CTCGTAGCCACTGTTGGTAAACAGGCAGCAGCAAAACTTGCAGAAAATATTGGCAAAATA 3836

Qy     421 ATGGGCGAAACGATTAAGAAGTTGGTGCCTAACGTGCTGAAACAGTTGGCGCAAAACGGC 480
           || ||  |||| | |   |  | | || ||  | || ||  | ||  |    |
Db    3837 ATAGGTAAAACCCTCACAGACCTTATACCAAAGTTTCTCAAGAATTTTTCTTCTCAACTG 3896

Qy     481 AGCAAACTCTTTACCCAGGGGATGCAACGTATTACTAGCGGTCTGGGTAATGTGGGTAGC 540
```

```
Db   3897 GACGATTTAATCACTAATGCTGTTGCCAGATTAAATAAATTTCTTGGTGCAGCGGGTGAT 3956

Qy    541 AAGATGGGCCTGCAAACGAATGCCTTAAGTAAAGAGCTGGTAGGTAATACCCTAAATAAA 600

Db   3957 GAAGT-----------------AATATCCAAACAAATTATTTCCACCCATTTAAACCAA 3998

Qy    601 GTGGCGTTGGGCATGGAAGTCACGAATACCGCAGCCCAGTCAGCCGGTGGTGTTGCCGAG 660

Db   3999 GCAGTTTTATTAGGAGAAAGTGTTAACTCTGCCACACAAGCGGGAGGAAGTGTCGCTTCT 4058

Qy    661 GGCGTATTTATTAAAAATGCCAGCGAGGCGCTTGCTGATTTTATGCTCGCCCGTTTTGCC 720

Db   4059 GCTGTTTTCCAGAACAGCGCGTCGACAAATCTAGCAGACCTGACATTATCGAAATATCAA 4118

Qy    721 ATGGATCAGATTCAGCAGTGGCTTAAACAATCCGTAGAAATATTTGGTGAAAACCAGAAG 780

Db   4119 GTTGAACAACTGTCAAAATATATCAGTGAAGCAATAGAAAAATTCGGCCAATTGCAGGAA 4178

Qy    781 GTAACGGCGGAACTGCAAAAAGCCATGTCTTCTGCGGTACAGCAAAATGCGGATGCTTCG 840

Db   4179 GTAATTGCAGATCTATTAGCCTCAATGTCCAACTCTCAGGCTAATAGAACTGATGTTGCA 4238

Qy    841 CGTTTTATTCTGCGCCAGAGTCGCGCATAAAAACTGCCAAAATAAAGGGAGAAAAAT 897

Db   4239 AAAGCAATTTTGCAACAAACTACTGCTTGATACAAATAAGGAGAATGTTATGGAAAT 4295
```

```
RESULT   8
SHFIPABCD
LOCUS       SHFIPABCD    5255 bp    DNA             BCT       19-MAR-1991
DEFINITION  S.flexneri ipaBCD region encoding invasion plasmid antigens.
ACCESSION   J04117
NID         g152740
VERSION     J04117.1  GI:152740
KEYWORDS    ipaA gene; ipaB gene; ipaC gene; ipaD; membrane protein.
SOURCE      S.flexneri (strain M90T-W) DNA, clone pHC17.
  ORGANISM  Shigella flexneri
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Shigella.
REFERENCE   1  (bases 1 to 5255)
  AUTHORS   Venkatesan,M.M., Buysse,J.M. and Kopecko,D.J.
  TITLE     Characterization of invasion plasmid antigen genes (ipaBCD) from
            Shigella flexneri
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 85, 9317-9321 (1988)
  MEDLINE   89057927
COMMENT     Draft entry and clean copy of sequence [1] kindly submitted by
            M.M.Venkatesan, 03-JAN-1989.
FEATURES             Location/Qualifiers
     source          1. .5255
                     /organism="Shigella flexneri"
                     /db_xref="taxon:623"
     CDS             606. .1073
                     /note="18 kD protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA26521.1"
                     /db_xref="PID:g152741"
                     /db_xref="GI:152741"
                     /translation="MSLNITENESISTAVIDAINSGATLKDINAIPDDMMDDIYSYAY
                     DFYNKGRIEEAEVFFRFLCIYDFYNVDYIMGLAAIYQIKEQFQQAADLYAVAFALGKN
                     DYTPVFHTGQCQLRLKAPLKAKECFELVIQHSNDEKLKIKAQSYLDAIQDIKE"
     CDS             1079. .2821
                     /note="IpaB protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA26522.1"
                     /db_xref="PID:g152742"
                     /db_xref="GI:152742"
                     /translation="MHNVSTTTTGFPLAKILTSTELGDNTIQAANDAANKLFSLTIAD
                     LTANQNINTTNAHSTSNILIPELKAPKSLNASSQLTLLIGNLIQILGEKSLTALTNKI
                     TAWKSQQQARQQKNLEFSDKINTLLSETEGLTRDYEKQINKLKNADSKIKDLENKINQ
                     IQTRLSNLDPESPEKKKLSREEIQLTIKKDAAVKDRTLIEQKTLSIHSKLTDKSMQLE
                     KEIDSFSAFSNTASAEQLSTQQKSLTGLASVTQLMATFIQLVGKNNEESLKNDLALFQ
                     SLQESRKTEMERKSDEYAAEVRKAEELNRVMGCVGKILGALLTIVSVVAAAFSGGASL
                     ALAAVGLALMVTDAIVQAATGNSFMEQALNPIMKAVIEPLIKLLSDAFTKMLEGLGVD
                     SKKAKMIGSILGAIAGALVLVAAVVLVATVGKQAAAKLAENIGKIIGKTLTDLIPKFL
                     KNFSSQLDDLITNAVARLNKFLGAAGDEVISKQIISTHLNQAVLLGESVNSATQAGGS
                     VASAVFQNSASTNLADLTLSKYQVEQLSKYISEAIEKFGQLQEVIADLLASMSNSQAN
                     RTDVAKAILQQTTA"
     CDS             2784. .3932
                     /note="IpaC protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA26523.1"
                     /db_xref="PID:g152743"
                     /db_xref="GI:152743"
                     /translation="MLQKQFCNKLLLDTNKENVMEIQNTKPTQTLYTDISTKQTQSSS
                     ETQKSQNYQQIAAHIPLNVGKNPVLTTTLNDDQLLKLSEQVQHDSEIIARLTDKKMKD
                     LSEMSHTLTPENTLDISSLSSNAVSLIISVAVLLSALRTAETKLGSQLSLIAFDATKS
                     AAENIVRQGLAALSSSITGAVTQVGITGIGAKKTHSGISDQKGALRKNLATAQSLEKE
                     LAGSKLGLNKQIDTNITSPQTNSSTKFLGKNKLAPDNISLSTEHKTSLSSPDISLQDK
                     IDTQRRTYELNTLSAQQKQNIGRATMETSAVAGNISTSGGRYASALEEEEQLISQASS
                     KQAEEASQVSKEASQATNQLIQKLLNIIDSINQSKNSAASQIAGNIRA"
     CDS             3983. .4981
                     /note="IpaD protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA26524.1"
                     /db_xref="PID:g152744"
                     /db_xref="GI:152744"
                     /translation="MNITTLTINSISTSSFSPNNTNGSSTETVNSDIKTTTSSHPVSSL
                     TMLNDTLHNIRTTNQALKKELSQKTLTKTSLEEIALHSSQISMDVNKSAQLLDILSRN
                     EYPINKDARELLHSAPKEAELDGDQMISHRELWAKIANSINDINEQYLKVYEHAVSSY
                     TQMYQDFSAVLSSLAGWISPGGNDGNSVKLQVNSLKKALEELKEKYKDKPLYPANNTV
                     SQEQANKWLTELGGTIGKVSQKNGGYVVSINMTPIDNMLKSLDNLGGNGEVVLDNAKY
                     QAWNAGFSAEDETMKNNLQTLVQKYSNANSIFDNLVKVLSSTISSCTDTDKLFLHF"
     CDS             4990. .>5255
                     /note="IpaA protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA26525.1"
                     /db_xref="PID:g551981"
                     /db_xref="GI:551981"
                     /translation="MHNVNNTQAPTFLYKATSPSSTEYSELKSKISDIHSSQTSLKTP
                     ASVSEKENFATSFNQKCLDFLFSSSGKEDVLRSIYSNSMNAYAK"
BASE COUNT     1950 a   1037 c    889 g   1379 t
ORIGIN
```

```
  Query Match             5.2%;  Score 179;  DB 1;  Length 5255;
  Best Local Similarity   51.2%;  Pred. No. 4.2e-35;
  Matches  459;  Conservative    0;  Mismatches  420;  Indels   18;  Gaps    1;

Qy       1 CGCAAAGCCGAGGAAACGAACCGCATTATGGGATGTATCGGGAAAGTCCTCGGCGCGCTG 60

Db    1970 CGTAAAGCAGAAGAACTCAACAGAGTAATGGGTTGTGTTGGGAAAATACTTGGGGCACTT 2029

Qy      61 CTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCT 120

Db    2030 TTAACTATCGTTAGTGTTGTTGCAGCAGCTTTTTCTGGAGGAGCCTCTCTAGCACTGGCA 2089

Qy     121 GCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCG 180

Db    2090 GCTGTTGGTTTAGCTCTTATGGTTACGGATGCTATAGTACAAGCAGCGACCGGCAATTCC 2149

Qy     181 TTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTG 240

Db    2150 TTCATGGAACAAGCCCTGAATCCGATCATGAAAGCAGTCATTGAACCCTTAATCAAACTC 2209

Qy     241 ATTGGCAAGGCGATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAG 300

Db    2210 CTTTCAGATGCATTTACAAAAATGCTCGAAGGCTTGGGCGTCGACTCGAAAAAAGCCAAA 2269

Qy     301 ATGGCCGGCAGCATTGTTGGTGCGATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTG 360
```

```
Db    2270 ATGATTGGCTCTATTCTGGGGGCAATCGCAGGCGCTCTTGTCCTAGTTGCAGCAGTCGTT 2329

Qy     361 GTGGTCGCAGTTGTCGGGAAAGGCGCGGCGGCGAAACTGGGTAACGCGCTGAGCAAAATG 420

Db    2330 CTCGTAGCCACTGTTGGTAAACAGGCAGCAGCAAAACTTGCAGAAAATATTGGCAAAATA 2389

Qy     421 ATGGGCGAAACGATTAAGAAGTTGGTGCCTAACGTGCTGAAACAGTTGGCGCAAAACGGC 480

Db    2390 ATAGGTAAAACCCTCACAGACCTTATACCAAAGTTTCTCAAGAATTTTTCTTCTCAACTG 2449

Qy     481 AGCAAACTCTTTACCCAGGGGATGCAACGTATTACTAGCGGTCTGGGTAATGTGGGTAGC 540

Db    2450 GACGATTTAATCACTAATGCTGTTGCCAGATTAAATAAATTTCTTGGTGCAGCGGGTGAT 2509

Qy     541 AAGATGGGCCTGCAAACGAATGCCTTAAGTAAAGAGCTGGTAGGTAATACCCTAAATAAA 600

Db    2510 GAAGT------------------AATATCCAAACAAATTATTTCCACCCATTTAAACCAA 2551

Qy     601 GTGGCGTTGGGCATGGAAGTCACGAATACCGCAGCCCAGTCAGCCGGTGGTGTTGCCGAG 660

Db    2552 GCAGTTTTATTAGGAGAAAGTGTTAACTCTGCCACACAAGCGGGAGGAAGTGTCGCTTCT 2611

Qy     661 GGCGTATTTATTAAAAATGCCAGCGAGGCGCTTGCTGATTTTATGCTCGCCCGTTTTGCC 720

Db    2612 GCTGTTTTCCAGAACAGCGCGTCGACAAATCTAGCAGACCTGACATTATCGAAATATCAA 2671

Qy     721 ATGGATCAGATTCAGCAGTGGCTTAAACAATCCGTAGAAATATTTGGTGAAAACCAGAAG 780

Db    2672 GTTGAACAACTGTCAAAATATATCAGTGAAGCAATAGAAAAATTCGGCCAATTGCAGGAA 2731

Qy     781 GTAACGGCGGAACTGCAAAAAGCCATGTCTTCTGCGGTACAGCAAAATGCGGATGCTTCG 840

Db    2732 GTAATTGCAGATCTATTAGCCTCAATGTCCAACTCTCAGGCTAATAGAACTGATGTTGCA 2791

Qy     841 CGTTTTATTCTGCGCCAGAGTCGCGCATAAAAACTGCCAAAATAAAGGGAGAAAAAT 897

Db    2792 AAAGCAATTTTGCAACAAACTACTGCTTGATACAAATAAGGAGAATGTTATGGAAAT 2848


RESULT   9
W10PPIBCA
LOCUS       W10PPIBCA     4761 bp    DNA             BCT       07-MAY-1996
DEFINITION  Plasmid pWR100 (from S.flexneri) antigen ippI, ipaB, ipaC genes,
            complete cds, and ippD gene, 3' end.
ACCESSION   M34849
NID         g155327
VERSION     M34849.1  GI:155327
KEYWORDS    antigen.
SOURCE      Plasmid pWR100 (clone: HB101/pHS[4108,5357].).
  ORGANISM  Plasmid pWR100
            plasmids.
REFERENCE   1  (bases 1 to 4761; 1 to 538; 3987 to 4761)
  AUTHORS   Baudry,B., Kaczorek,M. and Sansonetti,P.J.
  TITLE     Nucleotide sequence of the invasion plasmid antigen B and C genes
            (ipaB and ipaC) of Shigella flexneri
  JOURNAL   Microb. Pathog. 4 (5), 345-357 (1988)
  MEDLINE   89200844
COMMENT     Draft entry and computer-readable sequence for [1] kindly submitted
            by B.Baudry, 05-JUN-1990.
               Center for Vaccine Development
               University of Maryland at Baltimore
               10 South Pine Street
               Baltimore, MD 21201.
FEATURES             Location/Qualifiers
     source          1. .4761
                     /organism="Plasmid pWR100"
                     /plasmid="Plasmid pWR100"
                     /specific_host="M90T"
                     /db_xref="taxon:2661"
                     /clone="HB101/pHS[4108,5357]."
     CDS             <1. .549
                     /note="ORF"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA98422.1"
                     /db_xref="PID:g155328"
                     /db_xref="GI:155328"
                     /translation="KLVKKILSVFNLKQRFPQKNFGCPVNINKIRDSVIDKIKDSNSG
                     NQLFCWMSQERTTYVSSMINRSIDEMAIHNGVVLTSDNKKNIFAAIEKKFPDIKLDEK
                     SAQTSISHTALNEIASSGLRAKILKRYSSDMDLFNTQMKDLTNLVSSSVYDKIFNEST
                     KVLQIEISAEVLKAVYRQSNTN"
     -10_signal      556. .561
                     /gene="ippI"
     gene            556. .1073
                     /gene="ippI"
     RBS             595. .600
                     /gene="ippI"
     CDS             606. .1073
                     /gene="ippI"
                     /codon_start=1
                     /transl_table=11
                     /product="antigen"
                     /protein_id="AAA98423.1"
                     /db_xref="PID:g155329"
                     /db_xref="GI:155329"
                     /translation="MSLNITENESISTAVIDAINSGATLKDINAIPDDMMDDIYSYAY
                     DFYNKGRIEEAEVFFRFLCIYDFYNVDYIMGLAAIYQIKEQFQQAADLYAVAFALGKN
                     DYTPVFHTGQCQLRLKAPLKAKECFELVIQHSNDEKLKIKAQSYLDAIQDIKE"
     gene            1065. .2821
                     /gene="ipaB"
     RBS             1065. .1070
                     /gene="ipaB"
     CDS             1079. .2821
                     /gene="ipaB"
                     /codon_start=1
                     /transl_table=11
                     /product="antigen"
                     /protein_id="AAA98424.1"
                     /db_xref="PID:g155330"
                     /db_xref="GI:155330"
                     /translation="MHNVSTTTGFPLAKILTSTELGDNTIQAANDAANKLFSLTIAD
                     LTANQNINTTNAHSTSNILIPELKAPKSLNASSQLTLLIGNLIQILGEKSLTALTNKI
                     TAWKSQQQARQQKNLEFSDKINTLLSETEGLTRDYEKQINKLKNADSKIKDLENKINQ
                     IQTRLSELDPESPEKKKLSREEIQLTIKKDAAVKDRTLIEQKTLSIHSKLTDKSMQLE
                     KEIDSFSAFSNTASAEQLSTQQKSLTGLASVTQLMATFIQLVGKNNEESLKNDLALFQ
                     SLQESRKTEMERKSDEYAAEVRKAEELNRVMGCVGKILGALLTIVSVVAAAFSGGASL
                     ALAAVGLALMVTDAIVQAATGNSFMEQALNPIMKAVIEPLIKLLSDAFTKMLEGLGVD
                     SKKAKMIGSILGAIAGALVLVAAVVLVATVGKQAAAKLAENIGKIIGKTLTDLIPKFL
                     KNFSSQLDDLITNAVARLNKFLGAAGDEVISKQIISTHLNQAVLLGESVNSATQAGGS
                     VASAVFQNSASTNLADLTLSKYQVEQLSKYISEAIEKFGQLQEVIADLLASMSNSQAN
                     RTDVAKAILQQTTA"
     gene            2723. .3932
                     /gene="ipaC"
     -35_signal      2723. .2728
                     /gene="ipaC"
     -10_signal      2745. .2750
                     /gene="ipaC"
     RBS             2829. .2834
                     /gene="ipaC"
     CDS             2841. .3932
                     /gene="ipaC"
                     /codon_start=1
                     /transl_table=11
                     /product="antigen"
                     /protein_id="AAA98425.1"
                     /db_xref="PID:g155331"
                     /db_xref="GI:155331"
                     /translation="MEIQNTKPTQTLYTDISTKQTQSSSETQKSQNYQQIAAHIPLNV
                     GKNPVLTTTLNDDQLLKLSEQVQHDSEIIARLTDKKMKDLSEMSHILTPENTLDISSL
                     SSNAVSLIISVAVLLSALRTAETKLGSQLSLIAFDATKSAAENIVRQGLAALSSSITG
                     AVTQVGITGIGAKKTHSGISDQKGALRKNLATAQSLEKELAGSKLGLNKQIDTNITSP
                     QTNSSTKFLGKNKLAPDNISLSTEHKTSLSSPDISLQDKIDTQRRTIYELNTLSAQQKQ
                     NIGRATMETSAVAGNISTSGGRYASALEEEEQLISQASSKQAEEASQVSKEASQATNQ
                     LIQKLLNIIDSINQSKNSAASQIAGNIRA"
     RBS             3971. .3977
                     /gene="ipaD"
```

```
     gene            3971. .4761
                     /gene="ipaD"
     CDS             3983. .>4761
                     /gene="ipaD"
                     /codon_start=1
                     /transl_table=11
                     /product="antigen"
                     /protein_id="AAA98426.1"
                     /db_xref="PID:g552045"
                     /db_xref="GI:552045"
                     /translation="MNITTLTNSISTSSFSPNNTNGSSTETVNSDIKTTTSSHPVSSL
                     TMLNDTLHNIRTTNQALKKELSQKTLTKTSLEEIALHSSQISMDVNKSAQLLDILSRN
                     EYPINKDARELLHSAPKEAELDGDQMISHRELWAKIANSINDINEQYLKVYEHAVSSY
                     TQMYQDFSAVLSSLAGWISPGGNDGNSVKLQVNSLKKALEELKEKYKDKPLYPANNTV
                     SQEQANKWLTELGGTIGKVSQKNGGYVVSINMTPIDNMLKS"
BASE COUNT     1769 a    952 c    805 g   1235 t
ORIGIN
```

```
  Query Match             5.2%;  Score 179;  DB 1;  Length 4761;
  Best Local Similarity   51.2%;  Pred. No. 4.1e-35;
  Matches  459;  Conservative    0;  Mismatches  420;  Indels   18;  Gaps    1;

Qy       1 CGCAAAGCCGAGGAAACGAACCGCATTATGGGATGTATCGGGAAAGTCCTCGGCGCGCTG 60
           || ||||| || |||   ||| |  | ||||| ||| | |||||| | || || || ||
Db    1970 CGTAAAGCAGAAGAACTCAACAGAGTAATGGGTTGTGTTGGGAAAATACTTGGGGCACTT 2029

Qy      61 CTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCT 120
           |||| || || || ||||| || || | |||| | || || ||   ||| || |||||
Db    2030 TTAACTATCGTTAGTGTTGTTGCAGCAGCTTTTTCTGGAGGAGCCTCTCTAGCACTGGCA 2089

Qy     121 GCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCG 180
           || || ||  | ||  | ||||| | |||| || ||  | || ||||| ||    ||
Db    2090 GCTGTTGGTTTAGCTCTTATGGTTACGGATGCTATAGTACAAGCAGCGACCGGCAATTCC 2149

Qy     181 TTTATTCAGCAGGCGCTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTG 240
           || ||  | || || || || ||||| ||| |    ||  |  | || |||||  | ||
Db    2150 TTCATGGAACAAGCCCTGAATCCGATCATGAAAGCAGTCATTGAACCCTTAATCAAACTC 2209

Qy     241 ATTGGCAAGGCGATTACCAAAGCGCTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAG 300
           ||    | ||  |||| ||| |||  |||| ||||| || |||||||   |||||  ||  |
Db    2210 CTTTCAGATGCATTTACAAAAATGCTCGAAGGCTTGGGCGTCGACTCGAAAAAAGCCAAA 2269

Qy     301 ATGGCCGGCAGCATTGTTGGTGCGATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTG 360
           |||   |||  ||| | || || || |  | |||| ||| | | || || | |  | ||
Db    2270 ATGATTGGCTCIATTCTGGGGGCAATCGCAGGCGCTCTTGTCCTAGTTGCAGCAGTCGTT 2329

Qy     361 GTGGTCGCAGTTGTCGGGAAAGGCGCGGCGGCGAAACTGGGTAACGCGCTGAGCAAAATG 420
           | || ||   ||| || |||   || || || ||||| |   |     |  |||||||
Db    2330 CTCGTAGCCACTGTTGGTAAACAGGCAGCAGCAAAACTTGCAGAAAATATTGGCAAAATA 2389

Qy     421 ATGGGCGAAACGATTAAGAAGTTGGTGCCTAACGTGCTGAAACAGTTGGCGCAAAACGGC 480
           || ||  |||| | |   |  | || ||  | || ||  | ||  |     |
Db    2390 ATAGGTAAAACCCTCACAGACCTTATACCAAAGTTTCTCAAGAATTTTTCTTCTCAACTG 2449

Qy     481 AGCAAACTCTTTACCCAGGGGATGCAACGTATTACTAGCGGTCTGGGTAATGTGGGTAGC 540
             | |  | || | |     |  | | ||    ||| |||   | ||||
Db    2450 GACGATTTAATCACTAATGCTGTTGCCAGATTAAATAAATTTCTTGGTGCAGCGGGTGAT 2509

Qy     541 AAGATGGGCCTGCAAACGAATGCCTTAAGTAAAGAGCTGGTAGGTAATACCCTAAATAAA 600
           | |                    ||   ||| |  |  |     |       |||| ||
Db    2510 GAAGT------------------AATATCCAAACAAATTATTTCCACCCATTTAAACCAA 2551

Qy     601 GTGGCGTTGGGCATGGAAGTCACGAATACCGCAGCCCAGTCAGCCGGTGGTGTTGCCGAG 660
           |  |  ||     |||       ||  | ||  | ||  | |  ||  |||| ||
Db    2552 GCAGTTTTATTAGGAGAAAGTGTTAACTCTGCCACACAAGCGGGAGGAAGTGTCGCTTCT 2611

Qy     661 GGCGTATTTATTAAAAATGCCAGCGAGGCGCTTGCTGATTTTATGCTCGCCCGTTTTGCC 720
           |  || ||    || |  ||           || || ||  | |   |  |    | |
Db    2612 GCTGTTTTCCAGAACAGCGCGTCGACAAATCTAGCAGACCTGACATTATCGAAATATCAA 2671

Qy     721 ATGGATCAGATTCAGCAGTGGCTTAAACAATCCGTAGAAATATTTGGTGAAAACCAGAAG 780
           | || ||  |     | |   | |   || |  |||||| ||| ||  ||   ||| |
```

```
Db    2672 GTTGAACAACTGTCAAAATATATCAGTGAAGCAATAGAAAAATTCGGCCAATTGCAGGAA 2731

Qy     781 GTAACGGCGGAACTGCAAAAAGCCATGTCTTCTGCGGTACAGCAAAATGCGGATGCTTCG 840
           ||||  || || ||    |    | |||||      |          | |    | |||| | |
Db    2732 GTAATTGCAGATCTATTAGCCTCAATGTCCAACTCTCAGGCTAATAGAACTGATGTTGCA 2791

Qy     841 CGTTTTATTCTGCGCCAGAGTCGCGCATAAAAACTGCCAAAATAAAGGGAGAAAAAT 897
               ||| |||  || | |   || | | |      |  | || | |  ||||
Db    2792 AAAGCAATTTTGCAACAAACTACTGCTTGATACAAATAAGGAGAATGTTATGGAAAT 2848


RESULT  10
AC007249
LOCUS       AC007249   207466 bp    DNA             HTG       09-APR-1999
DEFINITION  Homo sapiens clone NH0320M02, WORKING DRAFT SEQUENCE, 5 unordered
            pieces.
ACCESSION   AC007249
NID         g4580499
VERSION     AC007249.2  GI:4580499
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 207466)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 207466)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-APR-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Apr 9, 1999 this sequence version replaced gi:4572717.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    12220: contig of 12220 bp in length
            *   12221    12238: gap of unknown length
            *   12239    30966: contig of 18728 bp in length
            *   30967    30984: gap of unknown length
            *   30985    50458: contig of 19474 bp in length
            *   50459    50476: gap of unknown length
            *   50477   105779: contig of 55303 bp in length
            *  105780   105797: gap of unknown length
            *  105798   207466: contig of 101669 bp in length.
FEATURES            Location/Qualifiers
     source         1. .207466
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="NH0320M02"
BASE COUNT    51723 a  53391 c  53578 g  48702 t     72 others
ORIGIN
```

```
  Query Match             2.1%;  Score 72.6;  DB 35;  Length 207466;
  Best Local Similarity   47.7%;  Pred. No. 8.7e-08;
  Matches  213;  Conservative    0;  Mismatches  234;  Indels    0;  Gaps    0;

Qy      25 ATTATGGGATGTATCGGGAAAGTCCTCGGCGCGCTGCTAACCATTGTCAGCGTTGTGGCC 84
           ||  || ||     | |||||||  |     |  |     |  |  |     ||| ||||
Db 135622 ATGGTGTGAATGTGCTGGAAAGTGCCTTCTGATCCCAGAGTCCCTCAACTCGTGGTGGTG 135681

Qy      85 GCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTG 144
           |  ||       |  | ||||    ||| |  ||  ||||| |   | ||||| ||||||
Db 135682 GTGGTGGCGGTGGCGGTGGCGGTGGTGGTGGCGGTGGCGGTAGTGGTGGCGGTGATGGTG 135741
```

```
Qy    145 GCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCGTTTATTCAGCAGGCGCTAAACCCG 204
Db 135742 GCGGTGGCAGTGGCGGTGGTGGCGGCGGTGGTGGTGATGGTGATGGTGGCGGTGATGGTG 135801

Qy    205 ATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTGATTGGCAAGGCGATTACCAAAGCG 264
Db 135802 GTGATGATGGTGATGGTGGTGGTGGTGGTGATGGTGATGGTGGTGGTGGTGGTGATGGTG 135861

Qy    265 CTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAGATGGCCGGCAGCATTGTTGGTGCG 324
Db 135862 GTGGTGGTGGTGGTGGTGGTGATGGTGGTGGTGGTGGTGGTGGTGGTGATGATGGTAGTG 135921

Qy    325 ATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTGGTGGTCGCAGTTGTCGGGAAAGGC 384
Db 135922 GTGGTGATGGTGGTGGTGGTGGTGGTGGTGATGGTGGTGGTGGTGGTGGTGGTGGTGGTG 135981

Qy    385 GCGGCGGCGAAACTGGGTAACGCGCTGAGCAAAATGATGGGCGAAACGATTAAGAAGTTG 444
Db 135982 ATGGTGGTGGTGGTGGTGATGGTGGTGGTGGTGGTGATGGTGGTGGTGGTGGTGGTGATG 136041

Qy    445 GTGCCTAACGTGCTGAAACAGTTGGCG 471
Db 136042 GTGGTGGTGGTGGTGGTGGTGGTGG 136068
```

RESULT 11
AC007009
LOCUS AC007009 158880 bp DNA HTG 05-MAR-1999
DEFINITION Homo sapiens clone NH0560C01, WORKING DRAFT SEQUENCE, 3 unordered pieces.
ACCESSION AC007009
NID g4337223
VERSION AC007009.1 GI:4337223
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 158880)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 158880)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (05-MAR-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
COMMENT * NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 6180: contig of 6180 bp in length
* 6181 6198: gap of unknown length
* 6199 40395: contig of 34197 bp in length
* 40396 40413: gap of unknown length
* 40414 158880: contig of 118467 bp in length.
FEATURES Location/Qualifiers
source 1. .158880
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NH0560C01"
BASE COUNT 48816 a 32181 c 31490 g 46357 t 36 others
ORIGIN

Query Match 2.1%; Score 72; DB 34; Length 158880;
Best Local Similarity 49.5%; Pred. No. 1.2e-07;
Matches 186; Conservative 0; Mismatches 190; Indels 0; Gaps 0;

```
Qy     25 ATTATGGGATGTATCGGGAAAGTCCTCGGCGCGCTGCTAACCATTGTCAGCGTTGTGGCC 84
Db    667 ATGGTGGTAGTGGTGGGGGTAGTGGTGATGAAGGAGATAATGGTGGTAGTGGTGGTGGAG 726

Qy     85 GCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTG 144
Db    727 GTGATGGTGATGGTGGCGGAGGAGGTGATGGTGGTAGTGGTGGGGGTAGTGGTGTTGAAG 786

Qy    145 GCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCGTTTATTCAGCAGGCGCTAAACCCG 204
Db    787 GAGATGGTGGTGGTGGTGGTGGAGGTGATGGTGATGGTGATAGTGGTGGGGGTAGTTGTG 846

Qy    205 ATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTGATTGGCAAGGCGATTACCAAAGCG 264
Db    847 ATGAAGGAGATGGTGGTTGTGGTGGTGGTGGTGATGGTGGAGGAGGTGATGGTGGTAGTG 906

Qy    265 CTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAGATGGCCGGCAGCATTGTTGGTGCG 324
Db    907 GTGGGGGTAGTGGTGATTAAGGAGATAATGGTGGTGGTGGTTGTGGTTGTTGTGGAGGAG 966

Qy    325 ATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTGGTGGTCGCAGTTGTCGGGAAAGGC 384
Db    967 ATGGTGGTTGTTGTGGAAGAGGTGATGGTGGTAGTGCTGGGGGTAGTGGTGGTGAAGGAG 1026

Qy    385 GCGGCGGCGAAACTGG 400
Db   1027 ATGGTGGTGGTGGTGG 1042
```

RESULT 12
AC004928/c
LOCUS AC004928 137137 bp DNA PRI 17-APR-1999
DEFINITION Homo sapiens PAC clone DJ0914M06 from 7q33-q35, complete sequence.
ACCESSION AC004928
NID g4587636
VERSION AC004928.2 GI:4587636
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 137137)
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
REFERENCE 2 (bases 1 to 137137)
AUTHORS Geisel,C. and Cloud,J.
TITLE The sequence of Homo sapiens PAC clone DJ0914M06
JOURNAL Unpublished (1999)
REFERENCE 3 (bases 1 to 137137)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (12-JUN-1998) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 137137)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (17-APR-1999) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Apr 19, 1999 this sequence version replaced gi:3213069.
SUBMITTED BY: WUGSC
Genome Sequencing Center
Department of Genetics
Washington University
St. Louis MO 63108, USA
http://genome.wustl.edu/gsc
mailto:sapiens@watson.wustl.edu

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping

clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see http://www.nhgri.nih.gov/DIR/GTB/CHR7 , send mailto:egreen@nhgri.nih.gov , or see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
This clone was derived from human PAC library RPCI-5, prepared by Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://bacpac.med.buffalo.edu) using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor.
The clone may be obtained either from Genome Systems, Inc. (http://www.genomesystems.com) or Research Genetics, Inc. (http://www.resgen.com); or from Pieter de Jong.
VECTOR: pCYPAC2
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is DJ0733H01, 200 bp overlap; the clone sequenced to the right is DJ0842K16, 200 bp overlap. Actual start of this clone is at base position 197 of DJ0914M06; actual end is at 136941 of DJ0914M06.

There is a tandem repeat within this sequence at base positions 79937 to 80547. The tandem consists of TC rich sequence and there was a forced join between two clones at base position 80183. Based on the HindIII digest and the EcoRV digest the assembly looks correct.

FEATURES	Location/Qualifiers
source	1. .137137
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="7"
	/map="7q33-q35"
	/clone_lib="RPCI-5"
	/clone="DJ0914M06"
repeat_region	80. .439
	/rpt_family="L2"
repeat_region	917. .958
	/rpt_family="AT_rich"
repeat_region	959. .1459
	/rpt_family="L1"
repeat_region	1460. .1651
	/rpt_family="MER1_type"
repeat_region	1652. .3086
	/rpt_family="L1"
repeat_region	3913. .3947
	/rpt_family="AT_rich"
repeat_region	5654. .5696
	/rpt_family="(TA)n"
repeat_region	6993. .7548
	/rpt_family="MER21_g"
repeat_region	8083. .8315
	/rpt_family="MIR"
repeat_region	8738. .8817
	/rpt_family="L2"
repeat_region	11249. .12279
	/rpt_family="Mariner"
repeat_region	12280. .12586
	/rpt_family="Alu"
repeat_region	12587. .12824
	/rpt_family="Mariner"
repeat_region	12975. .13291
	/rpt_family="Alu"
repeat_region	13536. .13600
	/rpt_family="Alu"
misc_feature	14686. .15096
	/note="match to EST M78298 (NID:g272614)"
STS	complement(14686. .15096)
	/db_xref="GI:3660601"
repeat_region	14909. .15105
	/rpt_family="MIR"
repeat_region	15980. .16638
	/rpt_family="L1"
repeat_region	16981. .17519
	/rpt_family="L2"
repeat_region	18243. .18385
	/rpt_family="L1"
misc_feature	18883. .19159
	/note="match to EST T05081 (NID:g315233)"
repeat_region	19476. .19522
	/rpt_family="MER2_type"
repeat_region	19523. .19901
	/rpt_family="MaLR"
repeat_region	19902. .20807
	/rpt_family="MER2_type"
repeat_region	20806. .20907
	/rpt_family="MER2_type"
repeat_region	21492. .21803
	/rpt_family="Alu"
repeat_region	23485. .23843
	/rpt_family="MaLR"
repeat_region	23941. .24048
	/rpt_family="MER1_type"
repeat_region	24852. .25345
	/rpt_family="L2"
STS	complement(25557. .25753)
	/db_xref="GI:3800469"
misc_feature	25559. .25753
	/note="match to EST AA076702 (NID:g1836190)"
repeat_region	26221. .26272
	/rpt_family="MIR"
misc_feature	26393. .26597
	/note="match to EST AA077210 (NID:g1836684)"
repeat_region	28261. .28342
	/rpt_family="L1"
repeat_region	28343. .28649
	/rpt_family="Alu"
repeat_region	28650. .28675
	/rpt_family="L1"
repeat_region	28756. .29288
	/rpt_family="L1"
misc_feature	29340. .29573
	/note="match to EST AA338345 (NID:g1990593)"
repeat_region	30247. .30336
	/rpt_family="purine-rich"
repeat_region	31667. .31726
	/rpt_family="A-rich"
repeat_region	32386. .32487
	/rpt_family="MIR"
repeat_region	32626. .32936
	/rpt_family="Alu"
repeat_region	33045. .33193
	/rpt_family="Alu"
repeat_region	33712. .33804
	/rpt_family="(CA)n"
repeat_region	35827. .36129
	/rpt_family="Alu"
misc_feature	36930. .37225
	/note="match to EST AA076689 (NID:g1836317)"
STS	36930. .37225
	/db_xref="GI:3800536"

```
     repeat_region   37217. .37318
                     /rpt_family="MIR"
     repeat_region   39267. .39559
                     /rpt_family="Alu"
     repeat_region   39627. .39730
                     /rpt_family="L2"
     repeat_region   39814. .39942
                     /rpt_family="L1"
     repeat_region   40542. .40566
                     /rpt_family="(TAA)n"
     repeat_region   40567. .41589
                     /rpt_family="L1"
     repeat_region   42368. .42564

  Query Match              2.1%;  Score 71.4;  DB 42;  Length 137137;
  Best Local Similarity   47.6%;  Pred. No. 1.7e-07;
  Matches  210;  Conservative    0;  Mismatches  231;  Indels    0;  Gaps    0;

Qy      37 ATCGGGAAAGTCCTCGGCGCGCTGCTAACCATTGTCAGCGTTGTGGCCGCTGTTTTTACC 96
           |||  |   ||  | || | | || |     | ||     | ||||  |  ||  |
Db   80443 ATCATGGTGGTAGTGGGGGTGGTGGTGGTGGTGGTGGTGATGGTGGTGGTGGTGATGGTG 80384

Qy      97 GGTGGGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTGGCCGATGAAATT 156
           |  | |  |    ||| | |||  | |||||   | | |||  || ||   |  |   |
Db   80383 GTGGTGCTGGTGGTGGTGGTGGTGGTGGTGATGGTGGTGGTGATGATGGTGGTGGTG 80324

Qy     157 GTGAAGGCGGCGACGGGAGTGTCGTTTATTCAGCAGGCGCTAAACCCGATTATGGAGCAT 216
           |||| || || ||  |||  | | ||   |  ||   | |    | |  ||| |
Db   80323 GTGATGGTGGTGATGGTGGTGGTGGTGATGATGATGGTAGTGATGGGGGTGGTGGTGGTG 80264

Qy     217 GTGCTGAAGCCGTTAATGGAGCTGATTGGCAAGGCGATTACCAAAGCGCTGGAAGGATTA 276
           ||  |||  |  | |  ||| | || | |     || | |          | |||      |
Db   80263 GTAGTGATGGTGGTGGTGGTGATGGTGGTGGTGGTGGTGGTGGTGATGGTGGTGATGGTG 80204

Qy     277 GGCGTCGATAAGAAAACGGCAGAGATGGCCGGCAGCATTGTTGGTGCGATTGTCGCCGCT 336
           |  ||  |  | ||  | || || |||||  |       | || |  |   |  | |  |
Db   80203 GTGGTGGTGATGATGATGGTAGTGATGGGGGTGGTGGTGGTGGTGGTAGTGATGGTGGTG 80144

Qy     337 ATTGCCATGGTGGCGGTCATTGTGGTGGTCGCAGTTGTCGGGAAAGGCGCGGCGGCGAAA 396
           | |  ||||||| ||| || |||||||| |  || || | ||     | || || |  |
Db   80143 GTGGTGATGGTGGTGGTGATGGTGGTGGTGGTGGTGGTGATGATGGTGGTGGTGGTA 80084

Qy     397 CTGGGTAACGCGCTGAGCAAAATGATGGGCGAAACGATTAAGAAGTTGGTGCCTAACGTG 456
           ||  |  | | ||       ||||||   | |  |||     | | |  ||       |||
Db   80083 GCGGTGATGGTGGTGGTGGTGATGATGATGGTAGTGATGGTAATGGTTATGATGGTGGTG 80024

Qy     457 CTGAAACAGTTGGCGCAAAAC 477
           ||     | |||  ||  | |
Db   80023 ATGGTGGTGGTGGTGCTGATC 80003


RESULT  13
AC007263
LOCUS       AC007263   168844 bp    DNA             HTG       24-MAY-1999
DEFINITION  Homo sapiens chromosome 14 clone BAC 79J20 map 14q31, WORKING DRAFT
            SEQUENCE, 5 ordered pieces.
ACCESSION   AC007263
NID         g4885690
VERSION     AC007263.3  GI:4885690
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 168844)
  AUTHORS   Rowen,L., Madan,A., Qin,S., Abbasi,N., Dors,M., Dickhoff,R.,
            James,R., Loretz,C., Lasky,S., Madan,A., Prescott,S., Ratcliffe,A.,
            Shaffer,T. and Hood,L.
  TITLE     Sequencing of human chromosome 14q31 region
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 168844)
  AUTHORS   Rowen,L., Madan,A., Qin,S., Abbasi,N., Dors,M., Dickhoff,R.,
            James,R., Loretz,C., Lasky,S., Madan,A., Prescott,S., Ratcliffe,A.,
            Shaffer,T. and Hood,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-APR-1999) Multimegabase Sequencing Center, University
            of Washington, PO BOX 357730, Seattle, WA 98195, USA
COMMENT     On May 24, 1999 this sequence version replaced gi:4680438.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved
            *       1    23725: contig of 23725 bp in length
            *                   gap of unknown length
            *   23726    34192: contig of 10467 bp in length
            *                   gap of unknown length
            *   34193    50560: contig of 16368 bp in length
            *                   gap of unknown length
            *   50561    78056: contig of 27496 bp in length
            *                   gap of unknown length
            *   78057   168844: contig of 90788 bp in length.
FEATURES            Location/Qualifiers
     source         1. .168844
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="14"
                    /clone="BAC 79J20"
                    /map="14q31"
                    /clone_lib="RPCI-11"
                    /note="This clone contains part of the gene for human
                    checkpoint suppressor 1 (U68723)"
BASE COUNT    42649 a  36845 c  39967 g  49381 t      2 others
ORIGIN


  Query Match              2.0%;  Score 70.2;  DB 35;  Length 168844;
  Best Local Similarity   48.4%;  Pred. No. 3.4e-07;
  Matches  195;  Conservative    0;  Mismatches  208;  Indels    0;  Gaps    0;

Qy      76 GTTGTGGCCGCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTGCG 135
           || ||||  |  |   |    |  | | |   |||  | |||  | |||||   | | |
Db   89547 GTGGTGGTGGTGGAGGTGGTGGTGGAGATGGTGGTGGTGGTGGTGGTGGTGGTGGTG 89606

Qy     136 GTAATGGTGGCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCGTTTATTCAGCAGGCG 195
           ||  ||||||  ||     | ||| ||| || |  ||   ||  | |  |   |  || |
Db   89607 GTGGTGGTGGTGGAGATGGTGGTGGAGGTGGTGGTGGAGATGGTGGTGGTGGTGGTGGTG 89666

Qy     196 CTAAACCCGATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTGATTGGCAAGGCGATT 255
           |     | |  |||||   ||| || ||  | |  ||||| || | |    || | |
Db   89667 GTGGTGGTGGTGGTGGAGATGGTGGTGGAGGTGGTGGTGGAGATGGTGGTGGTGGTGGTG 89726

Qy     256 ACCAAAGCGCTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAGATGGCCGGCAGCATT 315
           |  | | ||||  |   | |  ||||   |||    ||  | | |||  | |
Db   89727 GAGATGGTGGTGGAGGTGGTGGTGGTCGTAGAGATGGTGGTGGTGGTGGTCGTAGAG 89786

Qy     316 GTTGGTGCGATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTGGTGGTCGCAGTTGTC 375
           | | | | | || | |     |    |||||| | |  | |||||||| |  || |
Db   89787 ATGGTGGTGGTGGTGGTGGTGGAGGTGGTGGTGGAGATGGTGGTGGTGGTGGTCGTAGAG 89846

Qy     376 GGGAAAGGCGCGGCGGCGAAACTGGGTAACGCGCTGAGCAAAATGATGGGCGAAACGATT 435
           |   |  | || || | |  |||     | | ||      || |||  ||   | |
Db   89847 ATGGTGGTGGTGGTGGTGGAGATGGTGGTGGTGGTGGTGGTGGTGGAGGAGGAGGTG 89906

Qy     436 AAGAAGTTGGTGCCTAACGTGCTGAAACAGTTGGCGCAAAACG 478
           | || ||||||   | ||| ||     | |||  |  |  | |
Db   89907 GTGGAGCTGGTGCTGGAGGTGGTGGTGGTGGTGGTGGAGGAGG 89949


RESULT  14
AC000097/c
LOCUS       AC000097   162269 bp    DNA             PRI       01-DEC-1998
```

```
DEFINITION  Homo sapiens Chromosome 22q11.2 PAC Clone p201m18 In DGCR Region,
            complete sequence.
ACCESSION   AC000097
NID         g3935200
VERSION     AC000097.1  GI:3935200
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 162269)
  AUTHORS   Hua,A. and Roe,B.A.
  TITLE     Homo sapiens Chromosome 22q11.2 PAC Clone p201m18 In DGCR Region
  JOURNAL   Unpublished (1996)
REFERENCE   2  (bases 1 to 162269)
  AUTHORS   Budarf,M.L. and Emanuel,B.S.
  JOURNAL   Unpublished (1996)
REFERENCE   3  (bases 1 to 162269)
  AUTHORS   Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-1996) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   4  (bases 1 to 162269)
  AUTHORS   Hua,A., Budarf,M.L., Emanuel,B.S. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-FEB-1998) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   5  (bases 1 to 162269)
  AUTHORS   Hua,A., Budarf,M.L., Emanuel,B.S. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-FEB-1998) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   6  (bases 1 to 162269)
  AUTHORS   Hua,A., Budarf,M.L., Emanuel,B.S. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-NOV-1998) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   7  (bases 1 to 162269)
  AUTHORS   Hua,A., Budarf,M.L., Emanuel,B.S. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-DEC-1998) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On Dec 1, 1998 this sequence version replaced gi:3845359.
FEATURES             Location/Qualifiers
     source          1. .162269
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="22"
                     /map="22q11.2"
                     /clone="p201m18"
BASE COUNT    31852 a  46339 c  47075 g  37003 t
ORIGIN
```

```
  Query Match              2.0%;  Score 67.8;  DB 11;  Length 162269;
  Best Local Similarity   47.0%;  Pred. No. 1.4e-06;
  Matches  210;  Conservative    0;  Mismatches  237;  Indels    0;  Gaps    0;

Qy      25 ATTATGGGATGTATCGGGAAAGTCCTCGGCGCGCTGCTAACCATTGTCAGCGTTGTGGCC 84
           ||  |||       | | |  |||  | |  | | || |     |  |    || ||||
Db 111986 ATGGTGGTGGTGGTGGTGGTAGTGATGGTGGTGGTGGTGGTAGTGATGGTGGTGGTGGTG 111927

Qy      85 GCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTG 144
           |  ||  | |  |  | || |    ||| | |||  | | |||   | | |||  || ||
Db 111926 GTGGTAGTGATGGTGGTGGTGGTGGTGGTGGTAGTGATGGTGGTGGTGGTGATG 111867

Qy     145 GCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCGTTTATTCAGCAGGCGCTAAACCCG 204
           |  |  |   | |||| || |  |  ||  |||  | |  |   |  || | | |    |
```

```
Db 111866 GTGGTGGTGGTAGTGATGGTGATGGTGGTGGTGGTGATGGTGGTGGTGGTGGTGATGGTG 111807

Qy     205 ATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTGATTGGCAAGGCGATTACCAAAGCG 264
           |  ||  |    || ||  |  | | | | | || | |    ||  | |       || |
Db 111806 GTAGTGATGGTGATGGTGGTGGTGGTGGTAGTGATGGTGGTGGTGGTGGTGGTGGTAGTG 111747

Qy     265 CTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAGATGGCCGGCAGCATTGTTGGTGCG 324
           |||  |   | |  || | | |     ||  | | |||  |  |   | || |  | |
Db 111746 ATGGTGGTGGTGGTGGTAGTGATGGTGGTGGTGGTGGTGGTAGTGATGGTGGTGGTGGTG 111687

Qy     325 ATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTGGTGGTCGCAGTTGTCGGGAAAGGC 384
           | || |  |  || |   |||||| ||| || |||||||| | |||  | | |   |
Db 111686 GTGGTGGTAGTGATGGTGGTGGTGGTGGTGATGGTGGTGGTGGTAGTGATGGTGGTGGTG 111627

Qy     385 GCGGCGGCGAAACTGGGTAACGCGCTGAGCAAAATGATGGGCGAAACGATTAAGAAGTTG 444
           | || ||       |||      | | ||     | ||||||  |    | |  || | ||
Db 111626 GTGGTGGTAGTGATGGTGGTGGTGGTGGTGGTAGTGATGGTGGTGGTGGTGGTGATGGTG 111567

Qy     445 GTGCCTAACGTGCTGAAACAGTTGGCG 471
           |||      ||  |||    | |||
Db 111566 GTGGTGGTAGTAGTGATGGTGGTGGTG 111540
```

```
RESULT  15
AC006549/c
LOCUS       AC006549   165321 bp    DNA             HTG       18-MAY-1999
DEFINITION  Homo sapiens, WORKING DRAFT SEQUENCE, 8 unordered pieces.
ACCESSION   AC006549
NID         g4835949
VERSION     AC006549.17  GI:4835949
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 165321)
  AUTHORS   Hu,P., Chen,F., Dumanski,J. and Roe,B.A.
  TITLE     Homo sapiens Chromosome 22q11 PAC Clone p215k21 Distal To DGCR
            Region
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 165321)
  AUTHORS   Hu,P., Chen,F., Dumanski,J. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-FEB-1999) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On May 16, 1999 this sequence version replaced gi:4803977.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     4176: contig of 4176 bp in length
            *     4177     4245: gap of unknown length
            *     4246     6464: contig of 2219 bp in length
            *     6465     6533: gap of unknown length
            *     6534    13373: contig of 6840 bp in length
            *    13374    13442: gap of unknown length
            *    13443    24265: contig of 10823 bp in length
            *    24266    24334: gap of unknown length
            *    24335    36267: contig of 11933 bp in length
            *    36268    36335: gap of unknown length
            *    36336    52372: contig of 16037 bp in length
            *    52373    52440: gap of unknown length
            *    52441    68271: contig of 15831 bp in length
            *    68272    68339: gap of unknown length
            *    68340   165321: contig of 96982 bp in length.
FEATURES             Location/Qualifiers
     source          1. .165321
                     /organism="Homo sapiens"
```

```
                     /db_xref="taxon:9606"
BASE COUNT    32764 a  49257 c  48950 g  33869 t    481 others
ORIGIN


  Query Match             2.0%;  Score 67.8;  DB 35;  Length 165321;
  Best Local Similarity   47.0%;  Pred. No. 1.4e-06;
  Matches  210;  Conservative    0;  Mismatches  237;  Indels    0;  Gaps    0;

Qy      25 ATTATGGGATGTATCGGGAAAGTCCTCGGCGCGCTGCTAACCATTGTCAGCGTTGTGGCC 84
           ||  |||       | | |  |||  | |  | | || |     |  |    || ||||
Db 110322 ATGGTGGTGGTGGTGGTGGTAGTGATGGTGGTGGTGGTGGTAGTGATGGTGGTGGTGGTG 110263

Qy      85 GCTGTTTTTACCGGTGGGGCGAGTCTGGCGCTGGCTGCGGTGGGACTTGCGGTAATGGTG 144
           |  ||  | |  |  | || |    ||| | |||  | | |||   | | |||  || ||
Db 110262 GTGGTAGTGATGGTGGTGGTGGTGGTGGTGGTGGTAGTGATGGTGGTGGTGGTGGTGATG 110203

Qy     145 GCCGATGAAATTGTGAAGGCGGCGACGGGAGTGTCGTTTATTCAGCAGGCGCTAAACCCG 204
           |  |  |   | |||| || |  |  ||  |||  | |  |   |  || | | |    |
Db 110202 GTGGTGGTGGTAGTGATGGTGATGGTGGTGGTGGTGATGGTGGTGGTGGTGGTGATGGTG 110143

Qy     205 ATTATGGAGCATGTGCTGAAGCCGTTAATGGAGCTGATTGGCAAGGCGATTACCAAAGCG 264
           |  ||  |    || ||  |  | |  | | | || | |    || | |      || |
Db 110142 GTAGTGATGGTGATGGTGGTGGTGGTGGTAGTGATGGTGGTGGTGGTGGTGGTGGTAGTG 110083

Qy     265 CTGGAAGGATTAGGCGTCGATAAGAAAACGGCAGAGATGGCCGGCAGCATTGTTGGTGCG 324
           |||  |   | |  || |  | |     ||  | | |||  |  |   | || |  | |
Db 110082 ATGGTGGTGGTGGTGGTAGTGATGGTGGTGGTGGTGGTGGTAGTGATGGTGGTGGTGGTG 110023

Qy     325 ATTGTCGCCGCTATTGCCATGGTGGCGGTCATTGTGGTGGTCGCAGTTGTCGGGAAAGGC 384
           | || |  |  || |  |||||| ||| || |||||||| | |||  | | |   |
Db 110022 GTGGTGGTAGTGATGGTGGTGGTGGTGGTGATGGTGGTGGTGGTAGTGATGGTGGTGGTG 109963

Qy     385 GCGGCGGCGAAACTGGGTAACGCGCTGAGCAAAATGATGGGCGAAACGATTAAGAAGTTG 444
           | || ||      |||     | | ||      | ||||||  |    | |   || | ||
Db 109962 GTGGTGGTAGTGATGGTGGTGGTGGTGGTGGTAGTGATGGTGGTGGTGGTGGTGATGGTG 109903

Qy     445 GTGCCTAACGTGCTGAAACAGTTGGCG 471
           |||       ||| ||     | ||| |
Db 109902 GTGGTGGTAGTGATGGTGGTGGTGGTG 109876


Search completed: October  5, 1999, 18:07:56
Job time: 11919 sec
```

OM of: US-09-068-804-14 to: N_Geneseq_36:* out_format : pfs

Date: Oct 5, 1999 8:04 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09068804/runat_05101999_145135_10341/app_query.fasta.1
-DB=N_Geneseq_36 -QFMI=fastap -SUFFIX=rng -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=pam150
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=escore
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09068804
-NCPU=6 -ICPU=3 -WAIT -THREADS=1

Search information block:
Query: US-09-068-804-14
Query length: 513
Database: N_Geneseq_36:*
Database sequences: 311585
Database length: 125096042
Search time (sec): 156.120000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ..
N_Geneseq_36:T67036	+	2299.00	3897.09	1.8e-209	3390	! Salmonella secreted protein Ss
N_Geneseq_36:V30458_1	+	619.00	992.87	1.1e-47	110000	! Continuation (2 of 6) of V3
N_Geneseq_36:V30459_1	+	619.00	992.87	1.1e-47	110000	! Continuation (2 of 6) of V3
N_Geneseq_36:T30394	+	148.50	230.94	2.9e-05	1808	! Oligodendrocyte-myelin glycopr
N_Geneseq_36:Q50053	+	139.50	222.37	8.7e-05	924	! Decorin (N-terminal to half C-t
N_Geneseq_36:Q50046	+	139.50	221.55	9.7e-05	1002	! Mature decorin PT-65 coding se
N_Geneseq_36:T10741	+	139.50	221.31	0.0001	1026	! Human recombinant decorin DNA.
N_Geneseq_36:Q50052	+	135.50	216.39	0.0002	849	! Decorin (N-terminal to LRR10) c
N_Geneseq_36:X24012	+	143.00	213.65	0.0003	3949	! Human AOMF05 cDNA. Human G-prc
N_Geneseq_36:X24014	+	143.00	213.65	0.0003	3949	! Human AOMF05 cDNA. Human G-prc
N_Geneseq_36:X24013	+	143.00	213.64	0.0003	3950	! Human AOMF05 cDNA. Human G-prc
N_Geneseq_36:X24011	+	143.00	213.53	0.0003	3993	! Human AOMF05 cDNA. Human G-prc
N_Geneseq_36:T16517	+	139.50	207.07	0.0006	4192	! Collagen A1/decorin chimeric g
N_Geneseq_36:Q50051	+	127.00	203.67	0.0010	711	! Decorin (N-terminal to LRR8) cc
N_Geneseq_36:T08768	+	128.00	192.88	0.0038	2446	! Rat biglycan cDNA. Proteoglyca
N_Geneseq_36:V49536	+	132.00	189.51	0.0059	6703	! Adenylate cyclase gene. New ge
N_Geneseq_36:T47647	+	121.50	189.00	0.0063	1198	! Pineal gland specific gene-1 e
N_Geneseq_36:X52265	+	125.00	186.01	0.0093	2906	! Protein PRO331 cDNA clone DNA4
N_Geneseq_36:Q04491	+	120.50	184.41	0.0114	1593	! Sequence encoding human bone p
N_Geneseq_36:T42136	+	119.00	183.36	0.0130	1371	! I2C-3 gene encoding Fusarium w
N_Geneseq_36:X52247	+	120.00	181.52	0.0164	1947	! Protein PRO239 cDNA clone DNA3
N_Geneseq_36:Q04490	+	119.00	181.28	0.0170	1685	! Sequence encoding human bone p
N_Geneseq_36:T35106	+	113.50	181.08	0.0174	679	! Down-regulated senescence clone
N_Geneseq_36:V30919	+	120.00	179.47	0.0214	2384	! Human secreted protein AY304_1
N_Geneseq_36:X52267	+	119.00	174.49	0.0405	3296	! Protein PRO332 cDNA clone DNA4
N_Geneseq_36:T42135	+	120.50	173.98	0.0433	4465	! I2C-2 gene encoding Fusarium w
N_Geneseq_36:Q85594	+	123.50	173.92	0.0436	7452	! Human glycoprotein V gene. Pla
N_Geneseq_36:X52226	+	115.00	169.91	0.0730	2639	! Protein PRO220 cDNA clone DNA3
N_Geneseq_36:V80668	+	116.00	169.86	0.0734	3138	! Human DNAX toll-like receptor
N_Geneseq_36:T42137	+	113.00	169.16	0.0803	2027	! I2C-4 gene encoding Fusarium w
N_Geneseq_36:V04445	+	116.50	166.37	0.1148	4822	! Sequence used in detection met
N_Geneseq_36:V37227	+	106.50	164.05	0.1547	1122	! Bovine CD14 coding sequence. N
N_Geneseq_36:X32491	+	105.50	160.05	0.2584	1407	! Seq ID No:6 used in a method f
N_Geneseq_36:T31008	+	95.00	160.01	0.2596	240	! Erythropoietin-derived hypoxia
N_Geneseq_36:X52256	+	111.50	159.59	0.2741	4053	! Protein PRO326 cDNA clone DNA3
N_Geneseq_36:X52256	+	110.50	158.86	0.3009	3679	! Protein PRO293 cDNA clone DNA3
N_Geneseq_36:N82455	+	98.00	158.16	0.3291	478	! Sequence of the bioadhesive pre
N_Geneseq_36:Q89794	+	100.50	157.73	0.3480	761	! Leishmania donovania A2 gene OR
N_Geneseq_36:Q91450	+	107.00	155.36	0.4713	2880	! Tomato Cf-9 cDNA. Isolated pla
N_Geneseq_36:T06304	+	107.00	155.36	0.4713	2880	! Tomato pathogen resistance gen
N_Geneseq_36:Q85844	+	101.00	154.94	0.4977	1091	! Leishmania donovani A2 gene. I
N_Geneseq_36:Q47211	+	95.00	154.57	0.5219	411	! DNA polymerase III holoenzyme p
N_Geneseq_36:T98153	+	95.00	154.57	0.5219	411	! E. coli DNA polymerase III p
N_Geneseq_36:Q50049	+	95.00	154.20	0.5467	426	! Decorin (N-terminal to LRR4)
N_Geneseq_36:Q50050	+	96.50	153.87	0.5704	567	! Decorin (N-terminal to LRR6)

seq_name: N_Geneseq_36:T67036

seq_documentation_block:
ID T67036 standard; DNA; 3390 BP.
AC T67036;
DT 04-AUG-1997 (first entry)
DE Salmonella secreted protein SspH gene.
KW Salmonella secreted protein; SspH; bacterial-mediated endocytosis;
KW diagnosis; therapy; vaccine; attenuation; virulence; ss.
OS Salmonella typhimurium.
FH Key Location/Qualifiers
FT misc_difference 13. .14
FT /*tag= a
FT /note= "bases 13-14 are illegible in Fig 29"
FT misc_difference 37. .38
FT /*tag= b
FT /note= "bases 37-38 are illegible in Fig 29"
FT cds 1698. .3236
FT /*tag= c
FT /note= "(Claim 13) translated sequence does not
FT fully correspond to SspH sequence of Fig
FT 27"
PN WO9718225-A1.
PD 22-MAY-1997.
PF 14-NOV-1996; U18504.
PR 14-NOV-1995; US-006733.
PA (GEHO) GEN HOSPITAL CORP.
PI Miller SI;
DR WPI; 97-289217/26.
DR P-PSDB; W15274.
PT New isolated Salmonella secreted proteins and related genes - used
PT to develop products for the detection, treatment or prevention of
PT Salmonella infections
PS Claim 13; Fig 29; 95pp; English.
CC A DNA molecule (T67036) includes an open reading frame coding for
CC Salmonella secreted protein SspH (W15274) plus 5' and 3' flanking
CC regions. Ssp polypeptides (see also W15268-71) and nucleic acids
CC (see also T67028-31) can be used to develop products for the
CC detection, treatment or prevention of Salmonella infections.
SQ Sequence 3390 BP; 890 A; 812 C; 939 G; 745 T;

alignment_scores:
Quality: 2299.00 Length: 513
Ratio: 4.499 Gaps: 1
Percent Similarity: 99.610 Percent Identity: 99.610

alignment_block:
US-09-068-804-14 x T67036 ..

Align seg 1/1 to: T67036 from: 1 to: 3390

```
      1 MetArgAspCysLeuAsnAsnGlyAsnProValLeuAsnValGlyAlaSe 17
        |||||||||||||||||||||||||||||||||||||||||||||||||||
   1698 ATGCGTGATTGCCTGAATAACGGCAATCCAGTGCTTAACGTGGGAGCGTC 1747

     17 rGlyLeuThrThrLeuProAspArgLeuProProHisIleThrThrLeuV 34
        |||||||||||||||||||||||||||||||||||||||||||||||||||
   1748 AGGTCTTACCACCTTACCAGACCGTTTACCACCGCATATTACAACACTGG 1797

     34 alIleProAspAsnAsnLeuThrSerLeuProGluLeuProGluGlyLeu 50
        |||||||||||||||||||||||||||||||||||||||||||||||||||
   1798 TTATTCCTGATAATAATCTGACCAGCCTGCCGGAGTTGCCGGAAGGACTA 1847

     51 ArgGluLeuGluValSerGlyAsnLeuGlnLeuThrSerLeuProSerLe 67
        |||||||||||||||||||||||||||||||||||||||||||||||||||
   1848 CGGGAGCTGGAGGTCTCTGGTAACCTACAACTGACCAGCCTGCCATCGCT 1897

     67 uProGlnGlyLeuGlnLysLeuTrpAlaTyrAsnAsnTrpLeuAlaSerL 84
```

```
1898 GCCGCAGGGACTACAGAAGCTGTGGGCCTATAATAATTGGCTGGCCAGCC 1947

  84 euProThrLeuProProGlyLeuGlyAspLeuAlaValSerAsnAsnGln 100
1948 TGCCGACGTTGCCGCCAGGACTAGGGGATCTGGCGGTCTCTAATAACCAG 1997

 101 LeuThrSerLeuProGluMetProProAlaLeuArgGluLeuArgValSe 117
1998 CTGACCAGCCTGCCGGAGATGCCGCCAGCACTACGGGAGCTGAGGGTCTC 2047

 117 rGlyAsnAsnLeuThrSerLeuArgAlaLeuProSerGlyLeuGlnLysL 134
2048 TGGTAACAACCTGACCAG.CTGCGCGCGCTGCCGTCAGGACTACAGAAGC 2096

 134 euTrpAlaTyrAsnAsnArgLeuThrSerLeuProGluMetSerProGly 150
2097 TGTGGGCCTATAATAATCGGCTGACCAGCCTGCCGGAGATGTCGCCAGGA 2146

 151 LeuGlnGluLeuAspValSerHisAsnGlnLeuThrArgLeuProGlnSe 167
2147 CTACAGGAGCTGGATGTCTCTCATAACCAGCTGACCCGCCTGCCGCAAAG 2196

 167 rLeuThrGlyLeuSerSerAlaAlaArgValTyrLeuAspGlyAsnProL 184
2197 CCTCACGGGTCTGTCTTCAGCGGCACGCGTATATCTGGACGGGAATCCAC 2246

 184 euSerValArgThrArgAspArgLeuCysGlyHisHisTrpProPheArg 200
2247 TGTCTGTACGCACTCGTGACAGGCTCTGCGGACATCATTGGCCATTCAGG 2296

 201 HisGlnAspThrLeuArgTyrGlyGlyAlaPheArgProArgGluAlaAr 217
2297 CATCAGGATACACTTCCATATGGCGGGGCCTTCCGTCCCCGGGAAGCCCG 2346

 217 gAlaLeuHisLeuAlaValAlaAspTrpLeuThrSerAlaArgGluGlyG 234
2347 GGCACTGCACCTGGCGGTCGCTGACTGGCTGACCTCTGCACGGGAGGGGG 2396

 234 luAlaAlaGlnAlaAspArgTrpGlnAlaPheGlyLeuGluAspAsnAla 250
2397 AAGCGGCCCAGGCAGACAGATGGCAGGCGTTCGGACTGGAAGATAACGCC 2446

 251 AlaAlaPheSerLeuValLeuAspArgLeuArgGluThrGluAsnPheLy 267
2447 GCCGCCTTCAGCCTGGTCCTGGACAGACTGCGTGAGACGGAAAACTTCAA 2496

 267 sLysAspAlaGlyPheLysAlaGlnIleSerSerTrpLeuThrGlnLeuA 284
2497 AAAAGACGCGGGCTTTAAGGCACAGATATCATCCTGGCTGACACAACTGG 2546

 284 laGluAspAlaAlaLeuArgAlaLysThrPheAlaMetAlaThrGluAla 300
2547 CTGAAGATGCTGCGCTGAGAGCAAAAACCTTTGCCATGGCAACAGAGGCA 2596

 301 ThrSerThrCysGluAspArgValThrHisAlaLeuHisGlnMetAsnAs 317
2597 ACATCAACCTGCGAGGACCGGGTCACACATGCCCTGCACCAGATGAATAA 2646

 317 nValGlnLeuValHisAsnAlaGluLysGlyGluTyrAspAsnAsnLeuG 334
2647 CGTACAACTGGTACATAATGCAGAAAAAGGGGAATACGACAACAATCTCC 2696

 334 lnGlyLeuValSerThrGlyArgGluMetPheArgLeuAlaThrLeuGlu 350
2697 AGGGGCTGGTTTCCACGGGGCGTGAGATGTTCCGCCTGGCAACACTGGAA 2746

 351 GlnIleAla.ArgGluLysAlaGlyThrLeuAlaLeuValAspAspValG 367
2747 CAGATTGCCCCGGGAAAAAGCCGGAACACTGGCTTTAGTCGATGACGTTG 2796

 367 luValTyrLeuAlaPheGlnAsnLysLeuLysGluSerLeuGluLeuThr 383
```

```
2797 AGGTCTATCTGGCGTTCCAGAATAAGCTGAAGGAATCACTTGAGCTGACC 2846

 384 SerValThrSerGluMetArgPhePheAspValSerGlyValThrValSe 400
2847 AGCGTGACGTCAGAAATGCGTTTCTTTGACGTTTCCGGCGTGACGGTTTC 2896

 400 rAspLeuGlnAlaAlaAspValGlnValLysThrAlaGluAsnSerGlyP 417
2897 AGACCTTCAGGCTGCGGACGTTCAGGTGAAAACCGCTGAAAACAGCGGGT 2946

 417 heSerLysTrpIleLeuGlnTrpGlyProLeuHisSerValLeuGluArg 433
2947 TCAGTAAATGGATACTGCAGTGGGGGCCGTTACACAGCGTGCTGGAACGC 2996

 434 LysValProGluArgPheAsnAlaLeuArgGluLysGlnIleSerAspTy 450
2997 AAAGTGCCGGAACGCTTTAACGCGCTTCGTGAAAAGCAAATATCGGATTA 3046

 450 rGluAspThrTyrArgLysLeuTyrAspGluValLeuLysSerSerGlyL 467
3047 TGAAGACACGTACCGGAAGCTGTATGACGAAGTGCTGAAATCGTCCGGGC 3096

 467 euValAspAspThrAspAlaGluArgThrIleGlyValSerAlaMetAsp 483
3097 TGGTCGACGATACCGATGCAGAACGTACTATCGGAGTAAGTGCGATGGAT 3146

 484 SerAlaLysLysGluPheLeuAspGlyLeuArgAlaLeuValAspGluVa 500
3147 AGTGCGAAAAAAGAATTTCTGGATGGCCTGCGCGCTCTTGTGGATGAGGT 3196

 500 lLeuGlySerTyrLeuThrAlaArgTrpArgLeuAsn 512
3197 GCTGGGTAGCTATCTGACAGCCCGGTGGCGTCTTAAC 3233
```

seq_name: N_Geneseq_36:V30458_1

seq_documentation_block:
Continuation (2 of 6) of V30458 from base 100001 (Rhizobium species plasmid pNGR234a.
WP Sequence split into 6 fragments LOCUS V30458 Accession V30458

	Fragment Name	Begin	End
WP	V30458_0	1	110000
WP	V30458_1	100001	210000
WP	V30458_2	200001	310000
WP	V30458_3	300001	410000
WP	V30458_4	400001	510000
WP	V30458_5	500001	534720

alignment_scores:
Quality: 619.00 Length: 444
Ratio: 2.063 Gaps: 12
Percent Similarity: 67.568 Percent Identity: 37.838

alignment_block:
US-09-068-804-14 x V30458_1 ..

Align seg 1/1 to: V30458_1 from: 1 to: 110000

```
  59 LeuGlnLeuThrSerLeuPro...SerLeuProGlnGlyLeuGlnLysLe 74
25086 CTATCCCTGACCCCCTTGTCCATCCCCCTCGTATCCGGGCTCCGGCGCCT 25135

  74 uTrpAlaTyrAsnAsnTrpLeuAlaSerLeuPro...ThrLeuProProG 90
25136 GAACGTCAACAACAATCAGCTAGGGGATCTGCCAGACACCCTCCCGGGAA 25185

  90 lyLeuGlyAspLeuAlaValSerAsnAsnGlnLeuThrSerLeuProGlu 106
25186 CGCTTCTGGAGCTGGAGGCCAGCGAGAACCGGCTGACCCGCCTGCCCGAC 25235

 107 MetProProAlaLeuArgGluLeuArgValSerGlyAsnAsnLeuThrSe 123
25236 CTTCCTGCCGGGCTCCAGAGGCTGAACGTCGAAAACAACCGGCTGACCAA 25285
```

```
  123 rLeuArg...AlaLeuProSerGlyLeuGlnLysLeuTrpAlaTyrAsnA 139
25286 TCTCCCCGAGCCCCTCCCGGCCGCGCTCGAGTGGCTGGGTGCTGGCTACA 25335

  139 snArgLeuThrSerLeuProGluMet...SerProGlyLeuGlnGluLeu 154
25336 ACCAGCTAACCCGCCTGCCCGAGATGATTCCCCCCGAGCTTATATGGCTG 25385

  155 AspValSerHisAsnGlnLeuThrArgLeuProGlnSer...LeuThrGl 170
25386 GGCGCCAGAAACAATCAGCTGACCAGCGTGCCCGAAAGCCTGCTGACGCA 25435

  170 yLeuSerSerAlaAlaArgValTyrLeuAspGlyAsnProLeuSerValA 187
25436 GCTAGGCCAATGGTCTAGCATTGATTTGGAAAATAATCCGCTGCCGCATG 25485

  187 rgThrArgAspArgLeuCysGlyHisHisTrpProPheArgHisGlnAsp 203
25486 GGGTACAAACAAACTTGGTGACA.................GCCATGCAT 25517

  204 ThrLeuArgTyrGlyGlyAla......PheArgProArgGluAla..... 216
25518 GCCGCGGGTTACGCAGGACCGCAGATCTTTTTGCCTATGGGACCGGTGGA 25567

  217 .........ArgAlaLeuHisLeuAlaValAlaAspTrpLeuThrSerA 230
25568 ACTTGCACGACGGCCCCTTCACGAGGTTGTCGCGGATTGGCTC....... 25610

  230 laArgGluGlyGluAlaAlaGlnAlaAspArgTrpGlnAlaPheGlyLeu 246
25611 .....GAGGGGGATTTGGAAACGGTGGCCGCCTGGCGGGGCTTCGCCAAC 25655

  247 GluAspAsnAlaAlaAlaPheSerLeuValLeuAspArgLeuArgGluTh 263
25656 GAACAAGGTGCCCGGGATTATGCACATTTCCTCGACCGGCTCCGTACCAC 25705

  263 rGluAsnPheLysLysAspAlaGlyPheLysAlaGlnIleSerSerTrpL 280
25706 CGTGAACTATGGCAATGATGCG...TTCCGGCAGGCGGTGGCCATAGGTC 25752

  280 euThrGlnLeuAlaGluAspAlaAlaLeuArgAlaLysThrPheAlaMet 296
25753 TGCGGCAGGCAGTGGCCAGGCCGCAATTGCGCGCACAGTATTTTGAGCAG 25802

  297 AlaThrGluAlaThrSerThrCysGluAspArgValThrHisAlaLeuHi 313
25803 GCTTCCGGAGCTAGCGATAGCTGTGAGGATCGCATTACTTTGACCTGGAA 25852

  313 sGlnMetAsnAsnValGlnLeuValHisAsnAlaGluLysGlyGluTyrA 330
25853 TGGAATGCAGACCGCACTTCTGATCGCTGATGTAGAGGACGGGGTCTATG 25902

  330 spAsnAsnLeuGlnGlyLeuValSerThrGlyArgGluMetPheArgLeu 346
25903 ACGGGTCGCTCCACCAGCTTCTCCAGCACGGCCGTGTCATGTTCCGTTTG 25952

  347 AlaThrLeuGluGlnIleAlaArgGluLysAlaGlyThrLeu........ 360
25953 GAGGCACTGGACGGGATCGCGCGCGAGACGGTCAACTCGCTTCGACGTAC 26002

  361 .........AlaLeuValAspAspValGluValTyrLeuAlaPheGlnA 374
26003 CGATCCGGACGCCGACATCGACGAGATCGAGGTCTATCTTGCCTACCAGA 26052

  374 snLysLeuLysGluSerLeuGluLeuThrSerValThrSerGluMetArg 390
26053 CACAGCTGCGCGACACCCTGGAGCTGCGGCACGTCGCCCCAGACATGCGC 26102

  391 PhePheAspValSerGlyValThrValSerAspLeuGlnAlaAlaAspVa 407
26103 TTCTTGAACGTCTCTCACGTTACCGAGGAGGACGTTGCCAGAGCTGCCAG 26152

  407 lGlnValLysThrAlaGluAsnSerGlyPheSerLysTrpIle...LeuG 423
26153 TTCCGTGCGGGAACTTGAGGCGAGGGGATTCGGCGAGTATGTGGCAACGC 26202

  423 lnTrpGlyProLeuHisSerValLeuGluArgLysValProGluArgPhe 439
26203 GCTGGCAACCTTGGGAAAGAGTCATGAGGCGCATCGCCCCCGCAAGCCAT 26252

  440 AsnAlaLeuArgGluLysGlnIleSerAspTyrGluAspThrTyrArgLy 456
26253 GCAGCGATGCAGGAGCAGCTCATCGAAGCGATGGGAGAGGAGTTCAGAAG 26302

  456 sLeuTyrAspGluValLeuLysSerSerGlyLeuValAspAspThrAspA 473
26303 CCGCTTGGACGAAAAGCTCGCCGAGCACGGCCTGACGGGTGATGCCGATG 26352

  473 laGluArgThrIleGlyValSerAlaMetAsp 483
26353 CCGAACGGGTATTCGGAGCCGAGATCCTCAAC 26384
```

seq_name: N_Geneseq_36:V30459_1

seq_documentation_block:
Continuation (2 of 6) of V30459 from base 100001 (Rhizobium species symbiotic plasmid
WP Sequence split into 6 fragments LOCUS V30459 Accession V30459

	Fragment Name	Begin	End
WP	V30459_0	1	110000
WP	V30459_1	100001	210000
WP	V30459_2	200001	310000
WP	V30459_3	300001	410000
WP	V30459_4	400001	510000
WP	V30459_5	500001	536165

alignment_scores:
Quality: 619.00 Length: 444
Ratio: 2.063 Gaps: 12
Percent Similarity: 67.568 Percent Identity: 37.838

alignment_block:
US-09-068-804-14 x V30459_1 ..

Align seg 1/1 to: V30459_1 from: 1 to: 110000

```
   59 LeuGlnLeuThrSerLeuPro...SerLeuProGlnGlyLeuGlnLysLe 74
25086 CTATCCCTGACCCCCTTGTCCATCCCCCTCGTATCCGGGCTCCGGCGCCT 25135

   74 uTrpAlaTyrAsnAsnTrpLeuAlaSerLeuPro...ThrLeuProProG 90
25136 GAACGTCAACAACAATCAGCTAGGGGATCTGCCAGACACCCTCCCGGGAA 25185

   90 lyLeuGlyAspLeuAlaValSerAsnAsnGlnLeuThrSerLeuProGlu 106
25186 CGCTTCTGGAGCTGGAGGCCAGCGAGAACCGGCTGACCCGCCTGCCCGAC 25235

  107 MetProProAlaLeuArgGluLeuArgValSerGlyAsnAsnLeuThrSe 123
25236 CTTCCTGCCGGGCTCCAGAGGCTGAACGTCGAAAACAACCGGCTGACCAA 25285

  123 rLeuArg...AlaLeuProSerGlyLeuGlnLysLeuTrpAlaTyrAsnA 139
25286 TCTCCCCGAGCCCCTCCCGGCCGCGCTCGAGTGGCTGGGTGCTGGCTACA 25335

  139 snArgLeuThrSerLeuProGluMet...SerProGlyLeuGlnGluLeu 154
25336 ACCAGCTAACCCGCCTGCCCGAGATGATTCCCCCCGAGCTTATATGGCTG 25385

  155 AspValSerHisAsnGlnLeuThrArgLeuProGlnSer...LeuThrGl 170
25386 GGCGCCAGAAACAATCAGCTGACCAGCGTGCCCGAAAGCCTGCTGACGCA 25435

  170 yLeuSerSerAlaAlaArgValTyrLeuAspGlyAsnProLeuSerValA 187
```

```
25436 GCTAGGCCAATGGTCTAGCATTGATTTGGAAAATAATCCGCTGCCGCATG 25485

  187 rgThrArgAspArgLeuCysGlyHisHisTrpProPheArgHisGlnAsp 203
25486 GGGTACAAACAAACTTGGTGACA.................GCCATGCAT 25517

  204 ThrLeuArgTyrGlyGlyAla......PheArgProArgGluAla..... 216
25518 GCCGCGGGTTACGCAGGACCGCAGATCTTTTTGCCTATGGGACCGGTGGA 25567

  217 .........ArgAlaLeuHisLeuAlaValAlaAspTrpLeuThrSerA 230
25568 ACTTGCACGACGGCCCCTTCACGAGGTTGTCGCGGATTGGCTC....... 25610

  230 laArgGluGlyGluAlaAlaGlnAlaAspArgTrpGlnAlaPheGlyLeu 246
25611 .....GAGGGGGATTTGGAAACGGTGGCCGCCTGGCGGGGCTTCGCCAAC 25655

  247 GluAspAsnAlaAlaAlaPheSerLeuValLeuAspArgLeuArgGluTh 263
25656 GAACAAGGTGCCCGGGATTATGCACATTTCCTCGACCGGCTCCGTACCAC 25705

  263 rGluAsnPheLysLysAspAlaGlyPheLysAlaGlnIleSerSerTrpL 280
25706 CGTGAACTATGGCAATGATGCG...TTCCGGCAGGCGGTGGCCATAGGTC 25752

  280 euThrGlnLeuAlaGluAspAlaAlaLeuArgAlaLysThrPheAlaMet 296
25753 TGCGGCAGGCAGTGGCCAGGCCGCAATTGCGCGCACAGTATTTTGAGCAG 25802

  297 AlaThrGluAlaThrSerThrCysGluAspArgValThrHisAlaLeuHi 313
25803 GCTTCCGGAGCTAGCGATAGCTGTGAGGATCGCATTACTTTGACCTGGAA 25852

  313 sGlnMetAsnAsnValGlnLeuValHisAsnAlaGluLysGlyGluTyrA 330
25853 TGGAATGCAGACCGCACTTCTGATCGCTGATGTAGAGGACGGGGTCTATG 25902

  330 spAsnAsnLeuGlnGlyLeuValSerThrGlyArgGluMetPheArgLeu 346
25903 ACGGGTCGCTCCACCAGCTTCTCCAGCACGGCCGTGTCATGTTCCGTTTG 25952

  347 AlaThrLeuGluGlnIleAlaArgGluLysAlaGlyThrLeu........ 360
25953 GAGGCACTGGACGGGATCGCGCGCGAGACGGTCAACTCGCTTCGACGTAC 26002

  361 .........AlaLeuValAspAspValGluValTyrLeuAlaPheGlnA 374
26003 CGATCCGGACGCCGACATCGACGAGATCGAGGTCTATCTTGCCTACCAGA 26052

  374 snLysLeuLysGluSerLeuGluLeuThrSerValThrSerGluMetArg 390
26053 CACAGCTGCGCGACACCCTGGAGCTGCGGCACGTCGCCCCAGACATGCGC 26102

  391 PhePheAspValSerGlyValThrValSerAspLeuGlnAlaAlaAspVa 407
26103 TTCTTGAACGTCTCTCACGTTACCGAGGAGGACGTTGCCAGAGCTGCCAG 26152

  407 lGlnValLysThrAlaGluAsnSerGlyPheSerLysTrpIle...LeuG 423
26153 TTCCGTGCGGGAACTTGAGGCGAGGGGATTCGGCGAGTATGTGGCAACGC 26202

  423 lnTrpGlyProLeuHisSerValLeuGluArgLysValProGluArgPhe 439
26203 GCTGGCAACCTTGGGAAAGAGTCATGAGGCGCATCGCCCCCGCAAGCCAT 26252

  440 AsnAlaLeuArgGluLysGlnIleSerAspTyrGluAspThrTyrArgLy 456
26253 GCAGCGATGCAGGAGCAGCTCATCGAAGCGATGGGAGAGGAGTTCAGAAG 26302

  456 sLeuTyrAspGluValLeuLysSerSerGlyLeuValAspAspThrAspA 473
26303 CCGCTTGGACGAAAAGCTCGCCGAGCACGGCCTGACGGGTGATGCCGATG 26352

  473 laGluArgThrIleGlyValSerAlaMetAsp 483
26353 CCGAACGGGTATTCGGAGCCGAGATCCTCAAC 26384
```

```
seq_name: N_Geneseq_36:T30394

seq_documentation_block:
ID   T30394 standard; cDNA; 1808 BP.
AC   T30394;
DT   15-SEP-1996  (first entry)
DE   Oligodendrocyte-myelin glycoprotein cDNA.
KW   Oligodendrocyte-myelin glycoprotein; OMsp; CR-LR protein;
KW   central nervous system; CNS; myelination; neurofibromatosis;
KW   diagnosis; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             105. .1406
FT                   /*tag=  a
FT   signal_peptide  105. .155
FT                   /*tag=  b
FT   mat_peptide     156. .1358
FT                   /*tag=  c
FT   signal_peptide  1359. .1403
FT                   /*tag=  d
FT                   /note= "putative C-terminal signal peptide"
FT   unsure          145
FT                   /*tag=  e
FT                   /note= "base 145 was G in one clone and A in
FT                   another clone, leading to either glycine
FT                   or aspartic acid at position 14 of the
FT                   encoded protein"
PN   US5532351-A.
PD   02-JUL-1996.
PF   12-JUL-1990; 551267.
PR   12-JUL-1990; US-551267.
PR   21-JAN-1994; US-183588.
PA   (ARCH-) ARCH DEV CORP.
PI   Stefansson K;
DR   WPI; 96-321181/32.
DR   P-PSDB; R98454.
PT   Nucleic acid encoding oligodendrocyte-myelin glycoprotein - useful
PT   for diagnosis of neurofibromatosis
PS   Claim 2; Fig 1; 37pp; English.
CC   A cDNA sequence (T30394) codes for oligodendrocyte-myelin
CC   glycoprotein, Omgp (R98454), a protein involved in myelination in
CC   the CNS.  It was derived from overlapping clones isolated from a human
CC   neonatal spinal cord cDNA library using polyclonal rat anti-OMgp
CC   antibodies, and by re-screening with partial clones.  The OMgp gene
CC   bands at q11.2 of chromosome 17 and maps within 6 kDa of the t(1;17)
CC   translocation breakpoint that cosegregates with neurofibromatosis in
CC   some families.  The gene can be used as a probe to allow identification
CC   of persons, tissues or cells (including gametes) carrying the
CC   neurofibromatosis gene.
SQ   Sequence   1808 BP;    627 A;    397 C;    280 G;    503 T;


alignment_scores:
           Quality:  148.50              Length:    225
             Ratio:  1.160                 Gaps:     10
 Percent Similarity:  56.889   Percent Identity: 26.667

alignment_block:
 US-09-068-804-14 x T30394   ..

 Align seg 1/1  to: T30394  from: 1  to: 1808

      12 LeuAsnValGlyAlaSerGlyLeuThrThrLeuProAspArgLeuProPr 28
     198 GTGGACTGTTCAGGCAGAAACTTGTCTACATTACCATCTGGACTGCAAGA 247

      28 oHisIleThrThrLeuValIleProAspAsnAsnLeuThrSerLeu...P 44
```

```
      :::|||:::   |||   ::::::   |||:::::::|||:::|||
  248 GAATATTATACATTTAAACCTGTCITATAACCACTTTACTGATCTGCATA 297

       44 roGluLeuProGlu......GlyLeuArgGluLeuGluValSerGlyAsn 58
          :::|||::::::      :::||||||   |||:::::|||:::|||
  298 ACCAGTTAACCCAATATACCAATCTGAGGACCCTGGACATTTCAAACAAC 347

       59 LeuGlnLeuThrSerLeuProSer...LeuProGlnGlyLeuGlnLysLe 74
          :::|||   |||||||||:::   ||||||::::::|||   :::::
  348 ...AGGCTTGAAAGCCTGCCTGCTCACTTACCTCGGTCTCTGTGGAACAT 394

       74 uTrpAlaTyrAsnAsn..................................... 79
          :   |||   ||||||
  395 GTCTGCTGCTAACAACAACATTAAACTTCTTGACAAATCTGATACTGCTT 444

       80 .....TrpLeuAlaSerLeuProThrLeuProProGlyLeuGlyAspLeu 94
             |||                              :::|||      |||
  445 ATCAGTGG.........................AATCTTAAATATCTG 467

       95 AlaValSerAsnAsnGlnLeuThrSerLeuProGluMetProProAlaLe 111
          :::||||||:::|||   |||      :::      :::      :::||
  468 GATGTTTCTAAGAACATGCTGGAAAAGGTTGTCCTCATTAAAAATACACT 517

      111 uArgGluLeuArgValSerGlyAsnAsnLeuThrSerLeuArgAlaLeuP 128
          ||||   |||   |||                                |||
  518 AAGAAGTCTCGAGGTT.............................CTCA 537

      128 roSerGlyLeuGlnLysLeuTrpAlaTyrAsnAsnArgLeuThrSerLeu 144
             :::   :::|||||||||                  |||
  538 ACCTCAGTAGTAACAAACTTTGG.................ACAGTTCCA 569

      145 ProGluMetSerProGlyLeuGlnGluLeuAspValSerHisAsnGlnLe 161
          :::::::|||::::::   |||:::   :::|||:::|||:::|||   ||
  570 ACCAACATGCCCTCCAAACTACATATCGTGGACCTGTCTAATAATTCTTT 619

      161 uThrArg...LeuProGlnSerLeuThrGlyLeuSerSerAlaAlaArgV 177
          ||||:::   ||||||   :::|||:::::::|||::::::   :::::::
  620 GACACAAATTCTTCCAGGTACATTAATAAACCTGACAAATCTCACACATC 669

      177 alTyrLeuAspGlyAsnProLeuSerValArgThrArgAspArg...... 191
          ::|||||:::::|||   ::::::      :::   :::
  670 TTTACCTGCACAACAATAAGTTCACATTCATTCCAGACCAATCTTTTGAC 719

      192 ..........................LeuCysGlyHisHisTrpProPh 199
                                   |||:::::::::::::|||:::
  720 CAACTCTTTCAGTTGCAAGAGATAACCCTTTACAATAACAGGTGGTCATG 769

      199 eArgHisGlnAspThrLeuArgTyr 207
             |||:::::::::::::   |||
  770 TGACCACAAACAAAACATTACTTAC 794
```

```
seq_name: N_Geneseq_36:Q50053

seq_documentation_block:
ID   Q50053 standard; DNA; 924 BP.
AC   Q50053;
DT   28-APR-1994  (first entry)
DE   Decorin (N-terminal to half C-terminal) coding sequence PT-78.
KW   leucine-rich repeat; proteoglycan; cell regulatory factor; MBP;
KW   fusion protein; maltose binding protein; tumour growth; inhibition;
KW   decorin; PG-II; PG-40; transforming growth factor-beta; TGF-beta; ds.
FH   Key             Location/Qualifiers
FT   cds             1. .918
FT                   /*tag=  a
FT                   /product= PT-78
FT                   /note= "encodes decorin N-terminal sequence,
FT                   all 10 leucine-rich regions and half the
FT                   C-terminal region"
PN   WO9320202-A.
PD   14-OCT-1993.
PF   02-APR-1993; U03171.
PR   03-APR-1992; US-865652.
PA   (LJOL-) LA JOLLA CANCER RES FOUND.
PI   Cardenas J,  Craig W,  Mullen DG,  Pierschbacher MD;
PI   Ruoslahti EI;
DR   WPI; 93-336910/42.
DR   P-PSDB; R42267.
PT   Active fragments of protein esp. decorin - with cell regulatory
PT   factor domain, useful for inhibiting cell regulatory factor
PT   activity
PS   Example 5; Page 49-50; 77pp; English.
CC   Active fragments of decorin (full-length coding sequence Q50046)
CC   were generated by PCR and fused to Maltose Binding Protein. The
CC   resulting fusion proteins were useful for inhibiting the activity of
CC   a cell regulatory factor, esp. TGF-beta, and hence for treating
CC   conditions associated with over-activity of the growth factor such
CC   as certain tumours.
SQ   Sequence   924 BP;    269 A;    225 C;    202 G;    228 T;
```

```
alignment_scores:
           Quality:  139.50              Length:    260
             Ratio:  0.962                Gaps:     14
 Percent Similarity:  55.769   Percent Identity: 25.769

alignment_block:
 US-09-068-804-14 x Q50053   ..

 Align seg 1/1  to: Q50053  from: 1  to: 924

      11 ValLeuAsnValGlyAlaSerGlyLeuThrThrLeuProAspArgLeuPr 27
         |||::::::   ::::::   ||||||   :::::|||      |||||
     106 GTGGTCCAGTGTTCTGATTTGGGTCTGGACAAAGTGCCAAAGGATCTTCC 155

      27 oProHisIleThrThrLeuValIleProAspAsnAsnLeuThrSerLeuP 44
         ||||:::::::|||   |||   ::::::::::|||::::::|||
     156 CCCTGACACAACTCTGCTAGACCTGCAAAACAACAAAATAACC....... 198

      44 roGluLeuProGluGly............LeuArgGluLeuGlu...... 54
           |||:::   :::|||            |||::::::|||:::
     199 ..GAAATCAAAGATGGAGACTTTAAGAACCTGAAGAACCTTCACGCATTG 246

      55 ......ValSerGlyAsnLeu........................... 59
               |||:::::::::::::
     247 ATTCTTGTCAACAATAAAATTAGCAAAGTTAGTCCTGGAGCATTTACACC 296

      60 .....................................GlnLeuThrSerL 64
                                            ||||||:::   |
     297 TTTGGTGAAGTTGGAACGACTTTATCTGTCCAAGAATCAGCTGAAGGAAT 346

      64 euPro...SerLeuProGlnGlyLeuGlnLysLeuTrpAlaTyrAsnAsn 79
         |||||      :::|||:::   ||||||   |||:::|||:::::|||
     347 TGCCAGAAAAAATGCCCAAAACTCTTCAGGAGCTGCGTGCCCATGAGAAT 396

      80 TrpLeuAlaSerLeuProThrLeuPro...ProGlyLeuGlyAspLeuAl 95
            ::::::   :::   :::::::::      ||||||:::::::::
     397 GAGATCACCAAAGTGCGAAAAGTTACTTTCAATGGACTGAACCAGATGAT 446

      95 aVal..................................................... 96
         |||
     447 TGTCATAGAACTGGGCACCAATCCGCTGAAGAGCTCAGGAATTGAAAATG 496

      97 .....................................SerAsnAsn 99
                                             :::::::::
     497 GGGCTTTCCAGGGAATGAAGAAGCTCTCCTACATCCGCATTGCTGATACC 546

     100 GlnLeuThrSerLeuProGlu...MetProProAlaLeuArgGluLeuAr 115
         ::::::||||||:::|||:::   :::||||||:::|||   ||||||::
     547 AATATCACCAGCATTCCTCAAGGTCTTCCTCCTTCCCTTACGGAATTACA 596

     115 gValSerGlyAsnAsnLeu..............ThrSerLeuArgAlaL 127
         :::::::||||||::::::               :::||||||::::::|
     597 TCTTGATGGCAACAAAATCAGCAGAGTTGATGCAGCTAGCCTGAAAGGAC 646

     127 euProSerGlyLeuGlnLysLeu...TrpAlaTyrAsnAsnArgLeuThr 142
         ||   ::::::|||   ||||||      ::::::|||:::      :::
```

```
     647 TG...AATAATTTGGCTAAGTTGGGATTGAGTTTCAACAGCATCTCTGCT 693

     143 ............SerLeuProGluMetSerProGlyLeuGlnGluLeuAs 155
                     ||||||:::::   :::|||   |||:::||||||::
     694 GTTGACAATGGCTCTCTGGCCAAC...ACGCCTCATCTGAGGGAGCTTCA 740

     155 pValSerHisAsnGlnLeuThrArgLeuProGlnSerLeuThrGlyLeuS 172
         ::::::::::|||:::|||||||||:::|||   :::|||:::
     741 CTTGGACAACAACAAGCTTACCAGAGTACCTGGTGGGCTGGCAGAGCATA 790

     172 erSerAlaAlaArgValTyrLeuAspGlyAsnProLeuSerValArgThr 188
                    |||||||||:::::|||   :::|||||
     791 AGTACATCCAGGTTGTCTACCTTCATAACAACAATATCTCTGTAGTTGGA 840

     189 ArgAspArgLeuCys......GlyHisHis 196
             :::   :::|||      ||||||:::
     841 TCAAGTGACTTCTGCCCACCTGGACACAAC 870

seq_name: N_Geneseq_36:Q50046

seq_documentation_block:
ID   Q50046 standard; DNA; 1002 BP.
AC   Q50046;
DT   28-APR-1994  (first entry)
DE   Mature decorin PT-65 coding sequence.
KW   leucine-rich repeat; proteoglycan; cell regulatory factor; MBP;
KW   fusion protein; maltose binding protein; tumour growth; inhibition;
KW   decorin; PG-II; PG-40; transforming growth factor-beta; TGF-beta; ds.
FH   Key             Location/Qualifiers
FT   cds             1. .996
FT                   /*tag=  a
FT                   /product= PT-65
FT                   /note= "mature decorin"
PN   WO9320202-A.
PD   14-OCT-1993.
PF   02-APR-1993; U03171.
PR   03-APR-1992; US-865652.
PA   (LJOL-) LA JOLLA CANCER RES FOUND.
PI   Cardenas J,  Craig W,  Mullen DG,  Pierschbacher MD;
PI   Ruoslahti EI;
DR   WPI; 93-336910/42.
DR   P-PSDB; R42260.
PT   Active fragments of protein esp. decorin - with cell regulatory
PT   factor domain, useful for inhibiting cell regulatory factor
PT   activity
PS   Claim 10; Page 36-38; 77pp; English.
CC   Active fragments of decorin (full-length coding sequence Q50046)
CC   were generated by PCR and fused to Maltose Binding Protein. The
CC   resulting fusion proteins were useful for inhibiting the activity of
CC   a cell regulatory factor, esp. TGF-beta, and hence for treating
CC   conditions associated with over-activity of the growth factor such
CC   as certain tumours.
SQ   Sequence   1002 BP;    289 A;    247 C;    219 G;    247 T;


alignment_scores:
            Quality:  139.50            Length:    260
              Ratio:  0.962               Gaps:     14
 Percent Similarity:  55.769   Percent Identity: 25.769

alignment_block:
 US-09-068-804-14 x Q50046   ..

 Align seg 1/1  to: Q50046  from: 1  to: 1002

      11 ValLeuAsnValGlyAlaSerGlyLeuThrThrLeuProAspArgLeuPr 27
         |||::::::   ::::::   ||||||   ::::::|||      |||||
     106 GTGGTCCAGTGTTCTGATTTGGGTCTGGACAAAGTGCCAAAGGATCTTCC 155

      27 oProHisIleThrThrLeuValIleProAspAsnAsnLeuThrSerLeuP 44
         ||||::::::|||   |||   ::::::::::|||::::::|||
     156 CCCTGACACAACTCTGCTAGACCTGCAAAACAACAAAATAACC....... 198

      44 roGluLeuProGluGly.....LeuArgGluLeuGlu..... 54
           |||:::   :::|||      |||::::::|||:::
     199 ..GAAATCAAAGATGGAGACTTTAAGAACCTGAAGAACCTTCACGCATTG 246

      55 ......ValSerGlyAsnLeu..... 59
              |||::::::::::::
     247 ATTCTTGTCAACAATAAAATTAGCAAAGTTAGTCCTGGAGCATTTACACC 296

      60 .....GlnLeuThrSerL 64
                        ||||||:::   |
     297 TTTGGTGAAGTTGGAACGACTTTATCTGTCCAAGAATCAGCTGAAGGAAT 346

      64 euPro...SerLeuProGlnGlyLeuGlnLysLeuTrpAlaTyrAsnAsn 79
         |||||      :::|||:::   ||||||   |||:::|||::::::|||
     347 TGCCAGAAAAAATGCCCAAAACTCTTCAGGAGCTGCGTGCCCATGAGAAT 396

      80 TrpLeuAlaSerLeuProThrLeuPro...ProGlyLeuGlyAspLeuAl 95
         ::::::   :::   :::::::::      ||||||:::::::::
     397 GAGATCACCAAAGTGCGAAAAGTTACTTTCAATGGACTGAACCAGATGAT 446

      95 aVal..... 96
          |||
     447 TGTCATAGAACTGGGCACCAATCCGCTGAAGAGCTCAGGAATTGAAAATG 496

      97 .....SerAsnAsn 99
                        :::::::::
     497 GGGCTTTCCAGGGAATGAAGAAGCTCTCCTACATCCGCATTGCTGATACC 546

     100 GlnLeuThrSerLeuProGlu...MetProProAlaLeuArgGluLeuAr 115
         ::::::||||||:::|||:::   :::||||||:::|||   ||||||::
     547 AATATCACCAGCATTCCTCAAGGTCTTCCTCCTTCCCTTACGGAATTACA 596

     115 gValSerGlyAsnAsnLeu.....ThrSerLeuArgAlaL 127
         :::::::||||||::::::          :::||||||::::::|
     597 TCTTGATGGCAACAAAATCAGCAGAGTTGATGCAGCTAGCCTGAAAGGAC 646

     127 euProSerGlyLeuGlnLysLeu...TrpAlaTyrAsnAsnArgLeuThr 142
         ||   :::::::|||   ||||||      ::::::|||:::      :::
     647 TG...AATAATTTGGCTAAGTTGGGATTGAGTTTCAACAGCATCTCTGCT 693

     143 ............SerLeuProGluMetSerProGlyLeuGlnGluLeuAs 155
                     ||||||:::::   :::|||   |||:::||||||::
     694 GTTGACAATGGCTCTCTGGCCAAC...ACGCCTCATCTGAGGGAGCTTCA 740

     155 pValSerHisAsnGlnLeuThrArgLeuProGlnSerLeuThrGlyLeuS 172
         ::::::::::|||:::|||||||||:::|||   :::|||:::
     741 CTTGGACAACAACAAGCTTACCAGAGTACCTGGTGGGCTGGCAGAGCATA 790

     172 erSerAlaAlaArgValTyrLeuAspGlyAsnProLeuSerValArgThr 188
                    |||||||||:::::|||   :::|||||
     791 AGTACATCCAGGTTGTCTACCTTCATAACAACAATATCTCTGTAGTTGGA 840

     189 ArgAspArgLeuCys......GlyHisHis 196
             :::   :::|||      ||||||:::
     841 TCAAGTGACTTCTGCCCACCTGGACACAAC 870

seq_name: N_Geneseq_36:T10741

seq_documentation_block:
ID   T10741 standard; cDNA; 1026 BP.
AC   T10741;
DT   20-AUG-1996  (first entry)
DE   Human recombinant decorin DNA.
KW   Decorin; PG-II; PG-40; proteoglycan; guanidinium ion; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   signal_peptide  1. .42
FT                   /*tag=  a
FT   mat_peptide     43. .1026
FT                   /*tag=  b
PN   WO9601842-A1.
PD   25-JAN-1996.
PF   07-JUL-1995; U08542.
```

```
PR   08-JUL-1994; US-272919.
PA   (LJOL-) LA JOLLA CANCER RES FOUND.
PI   Craig WS,  Harper JR,  Hernandez SD,  Kostel PJ,  Parker JR;
PI   Vedvick TS;
DR   WPI; 96-097586/10.
DR   P-PSDB; R89439.
PT   Purificn. of human recombinant decorin - using a strong anion
PT   exchange resin, a hydrophobic interaction chromatography resin and a
PT   strong anion exchange resin
PS   Disclosure; Fig 1A-D; 55pp; English.
CC   A cDNA sequence (T10741) is used for prodn. of human recombinant
CC   decorin (R89439) in CHO host cells.  It can be incorporated into
CC   vector pSV2.  DHFR-negative CHO-DG44 cells are then cotransfected
CC   with pSV2-decorin and pSV2dhfr.  Large-scale cultures can be
CC   performed using CHO cells attached to microcarrier beads.
CC   Recombinant decorin is purified from the cells using a 3-step
CC   chromatographic procedure.  It can be used for the highly sensitive
CC   detection of guanidinium ions (ppm range), partic. in protein-contg.
CC   solns. purified using GuHCl, and also has therapeutic applns.
SQ   Sequence   1026 BP;    293 A;    253 C;    226 G;    254 T;


alignment_scores:
            Quality:  139.50              Length:    260
              Ratio:  0.962                 Gaps:     14
 Percent Similarity:  55.769   Percent Identity: 25.769

alignment_block:
 US-09-068-804-14 x T10741   ..

 Align seg 1/1  to: T10741  from: 1  to: 1026

      11 ValLeuAsnValGlyAlaSerGlyLeuThrThrLeuProAspArgLeuPr 27
         |||::::::   ::::::   ||||||   ::::::|||      |||||
     142 GTGGTCCAGTGTTCTGATTTGGGTCTGGACAAAGTGCCAAAGGATCTTCC 191

      27 oProHisIleThrThrLeuValIleProAspAsnAsnLeuThrSerLeuP 44
         ||||::::::|||   |||   ::::::::::|||::::::|||
     192 CCCTGACACAACTCTGCTAGACCTGCAAAACAACAAAATAACC....... 234

      44 roGluLeuProGluGly.............LeuArgGluLeuGlu...... 54
           |||:::   :::|||              |||::::::|||:::
     235 ..GAAATCAAAGATGGAGACTTTAAGAACCTGAAGAACCTTCACGCATTG 282

      55 ......ValSerGlyAsnLeu........................... 59
              |||:::::::::::::
     283 ATTCTTGTCAACAATAAAATTAGCAAAGTTAGTCCTGGAGCATTTACACC 332

      60 .....................................GlnLeuThrSerL 64
                                             ||||||:::   |
     333 TTTGGTGAAGTTGGAACGACTTTATCTGTCCAAGAATCAGCTGAAGGAAT 382

      64 euPro...SerLeuProGlnGlyLeuGlnLysLeuTrpAlaTyrAsnAsn 79
         |||||      :::|||:::   ||||||   |||:::|||:::::|||
     383 TGCCAGAAAAAATGCCCAAAACTCTTCAGGAGCTGCGTGCCCATGAGAAT 432

      80 TrpLeuAlaSerLeuProThrLeuPro...ProGlyLeuGlyAspLeuAl 95
         ::::::   :::   :::::::::      ||||||:::::::::
     433 GAGATCACCAAAGTGCGAAAAGTTACTTTCAATGGACTGAACCAGATGAT 482

      95 aVal...................................................96
         |||
     483 TGTCATAGAACTGGGCACCAATCCGCTGAAGAGCTCAGGAATTGAAAATG 532

      97 .....................................SerAsnAsn 99
                                             :::::::::
     533 GGGCTTTCCAGGGAATGAAGAAGCTCTCCTACATCCGCATTGCTGACACC 582

     100 GlnLeuThrSerLeuProGlu...MetProProAlaLeuArgGluLeuAr 115
         ::::::||||||:::|||:::   :::|||||||:::|||   ||||||::
     583 AATATCACCAGCATTCCTCAAGGTCTTCCTCCTTCCCTTACGGAATTACA 632

     115 gValSerGlyAsnAsnLeu..............ThrSerLeuArgAlaL 127
         :::::::||||||::::::                ::::|||||||::::::|
     633 TCTTGATGGCAACAAAATCAGCAGAGTTGATGCAGCTAGCCTGAAAGGAC 682

     127 euProSerGlyLeuGlnLysLeu...TrpAlaTyrAsnAsnArgLeuThr 142
         ||   :::::::|||   ||||||      ::::::|||:::      :::
     683 TG...AATAATTTGGCTAAGTTGGGATTGAGTTTCAACAGCATCTCTGCT 729

     143 ............SerLeuProGluMetSerProGlyLeuGlnGluLeuAs 155
                     ||||||::::::   :::|||   |||::::::::::
     730 GTTGACAATGGCTCTCTGGCCAAC...ACGCCTCATCTGAGGGAGCTTCA 776

     155 pValSerHisAsnGlnLeuThrArgLeuProGlnSerLeuThrGlyLeuS 172
         ::::::::::|||:::|||||||||:::|||   :::|||:::
     777 CTTGGACAACAACAAGCTTACCAGAGTACCTGGTGGGCTGGCAGAGCATA 826

     172 erSerAlaAlaArgValTyrLeuAspGlyAsnProLeuSerValArgThr 188
                     |||||||||:::::|||   :::||||||
     827 AGTACATCCAGGTTGTCTACCTTCATAACAACAATATCTCTGTAGTTGGA 876

     189 ArgAspArgLeuCys......GlyHisHis 196
         :::   :::|||      |||||||:::
     877 TCAAGTGACTTCTGCCCACCTGGACACAAC 906

seq_name: N_Geneseq_36:Q50052

seq_documentation_block:
ID   Q50052 standard; DNA; 849 BP.
AC   Q50052;
DT   28-APR-1994  (first entry)
DE   Decorin (N-terminal to LRR10) coding sequence PT-77.
KW   leucine-rich repeat; proteoglycan; cell regulatory factor; MBP;
KW   fusion protein; maltose binding protein; tumour growth; inhibition;
KW   decorin; PG-II; PG-40; transforming growth factor-beta; TGF-beta; ds.
FH   Key             Location/Qualifiers
FT   cds             1. .843
FT                   /*tag= a
FT                   /product= PT-77
FT                   /note= "encodes decorin N-terminal sequence
FT                   and all 10 leucine-rich regions"
PN   WO9320202-A.
PD   14-OCT-1993.
PF   02-APR-1993; U03171.
PR   03-APR-1992; US-865652.
PA   (LJOL-) LA JOLLA CANCER RES FOUND.
PI   Cardenas J,  Craig W,  Mullen DG,  Pierschbacher MD;
PI   Ruoslahti EI;
DR   WPI; 93-336910/42.
DR   P-PSDB; R42266.
PT   Active fragments of protein esp. decorin - with cell regulatory
PT   factor domain, useful for inhibiting cell regulatory factor
PT   activity
PS   Example 5; Page 47-48; 77pp; English.
CC   Active fragments of decorin (full-length coding sequence Q50046)
CC   were generated by PCR and fused to Maltose Binding Protein. The
CC   resulting fusion proteins were useful for inhibiting the activity of
CC   a cell regulatory factor, esp. TGF-beta, and hence for treating
CC   conditions associated with over-activity of the growth factor such
CC   as certain tumours.
SQ   Sequence   849 BP;    250 A;    203 C;    187 G;    209 T;


alignment_scores:
            Quality:  135.50              Length:    248
              Ratio:  0.975                 Gaps:     13
 Percent Similarity:  56.048   Percent Identity: 25.806

alignment_block:
 US-09-068-804-14 x Q50052   ..

 Align seg 1/1  to: Q50052  from: 1  to: 849

      11 ValLeuAsnValGlyAlaSerGlyLeuThrThrLeuProAspArgLeuPr 27
         |||::::::   ::::::   ||||||   ::::::|||      |||||
```

```
   106 GTGGTCCAGTGTTCTGATTTGGGTCTGGACAAAGTGCCAAAGGATCTTCC 155

    27 oProHisIleThrThrLeuValIleProAspAsnAsnLeuThrSerLeuP 44
       ||||:::::::|||   |||   :::::::::|||:::::::|||
   156 CCCTGACACAACTCTGCTAGACCTGCAAAACAACAAAATAACC..... 198

    44 roGluLeuProGluGly...........LeuArgGluLeuGlu..... 54
         |||:::   :::|||            |||:::::::|||:::
   199 ..GAAATCAAAGATGGAGACTTTAAGAACCTGAAGAACCTTCACGCATTG 246

    55 ......ValSerGlyAsnLeu........................... 59
             |||:::::::::::::
   247 ATTCTTGTCAACAATAAAATTAGCAAAGTTAGTCCTGGAGCATTTACACC 296

    60 .....................................GlnLeuThrSerL 64
                                           ||||||:::   |
   297 TTTGGTGAAGTTGGAACGACTTTATCTGTCCAAGAATCAGCTGAAGGAAT 346

    64 euPro...SerLeuProGlnGlyLeuGlnLysLeuTrpAlaTyrAsnAsn 79
       |||||      :::|||:::   ||||||   |||:::|||:::::|||
   347 TGCCAGAAAAAATGCCCAAAACTCTTCAGGAGCTGCGTGCCCATGAGAAT 396

    80 TrpLeuAlaSerLeuProThrLeuPro...ProGlyLeuGlyAspLeuAl 95
         ::::::   :::   :::::::::      ||||||:::::::::
   397 GAGATCACCAAAGTGCGAAAAGTTACTTTCAATGGACTGAACCAGATGAT 446

    95 aVal..................................................  96
         |||
   447 TGTCATAGAACTGGGCACCAATCCGCTGAAGAGCTCAGGAATTGAAAATG 496

    97 .....................................SerAsnAsn 99
                                            :::::::::
   497 GGGCTTTCCAGGGAATGAAGAAGCTCTCCTACATCCGCATTGCTGATACC 546

   100 GlnLeuThrSerLeuProGlu...MetProProAlaLeuArgGluLeuAr 115
       ::::::||||||:::|||:::   :::||||||:::|||   ||||||::
   547 AATATCACCAGCATTCCTCAAGGTCTTCCTCCTTCCCTTACGGAATTACA 596

   115 gValSerGlyAsnAsnLeu..............ThrSerLeuArgAlaL 127
       :::::::||||||:::::                :::|||||||:::::|
   597 TCTTGATGGCAACAAAATCAGCAGAGTTGATGCAGCTAGCCTGAAAGGAC 646

   127 euProSerGlyLeuGlnLysLeu...TrpAlaTyrAsnAsnArgLeuThr 142
       ||   :::::::|||   ||||||      :::::|||:::      :::
   647 TG...AATAATTTGGCTAAGTTGGGATTGAGTTTCAACAGCATCTCTGCT 693

   143 ............SerLeuProGluMetSerProGlyLeuGlnGluLeuAs 155
                   ||||||::::::   :::|||   |||:::||||||::
   694 GTTGACAATGGCTCTCTGGCCAAC...ACGCCTCATCTGAGGGAGCTTCA 740

   155 pValSerHisAsnGlnLeuThrArgLeuProGlnSerLeuThrGlyLeuS 172
       ::::::::::|||:::|||||||||:::|||   :::|||:::
   741 CTTGGACAACAACAAGCTTACCAGAGTACCTGGTGGGCTGGCAGAGCATA 790

   172 erSerAlaAlaArgValTyrLeuAspGlyAsnProLeuSerVal 186
                 |||||||||:::::|||   :::|||||||
   791 AGTACATCCAGGTTGTCTACCTTCATAACAACAATATCTCTGTA 834

seq_name: N_Geneseq_36:X24012

seq_documentation_block:
ID   X24012 standard; cDNA; 3949 BP.
AC   X24012;
DT   29-JUN-1999  (first entry)
DE   Human AOMF05 cDNA.
KW   AOMF05; human; G-protein coupled glycoprotein hormone receptor; brain;
KW   anti-obesity; anti-diabetic; treatment; pancreatic disease; obesity;
KW   diabetes; endocrine system; skeletal muscle; spinal cord; placenta;
KW   transgenic animal; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             568. .3417
FT                   /*tag=  a
FT                   /product= "AOMF05"
PN   WO9915545-A1.
PD   01-APR-1999.
PF   24-SEP-1998; U20101.
PR   24-SEP-1997; US-059868.
PA   (MERI ) MERCK & CO INC.
PI   Abramovitz M, Liu Q, McDonald TP, ONeill GP, Wang R;
DR   WPI; 99-254689/21.
DR   P-PSDB; W93905.
PT   Human G-protein coupled glycoprotein hormone receptor AOMF05
PS   Claim 1; Fig 3A-F; 89pp; English.
CC   This invention describes the isolation of a novel human G-protein coupled
CC   glycoprotein hormone receptor, AOMF05 which has anti-obesity and
CC   anti-diabetic activity. The AOMF05 gene, receptor and modulators are
CC   useful for treatment of pancreatic disease, obesity and diabetes.
CC   Glycoprotein hormone receptors are important in the endocrine system
CC   and AOMF05 may be involved in development and function of the skeletal
CC   muscle, spinal cord, placenta and to a lesser extent, the brain. The
CC   transgenic animal may be useful for studying tissue and temporal
CC   specific expression or activity of the AOMF05 receptor, as well as
CC   for studying the ability of a variety of compounds to act as modulators
CC   of AOMF05 receptor activity.
SQ   Sequence   3949 BP;    1015 A;    912 C;    950 G;    1072 T;

alignment_scores:
          Quality:  143.00            Length:    206
            Ratio:  1.144               Gaps:      9
 Percent Similarity:  60.680   Percent Identity: 27.670

alignment_block:
 US-09-068-804-14 x X24012   ..

 Align seg 1/1  to: X24012  from: 1  to: 3949

      4 CysLeuAsnAsnGlyAsnProValLeuAsnValGlyAlaSerGlyLeuTh 20
        |||      :::|||:::      ::::::   ::::::   ||||||||
    664 TGCAGCTGCGACGGCGACCGTCGGGTGGACTGCTCCGGGAAGGGGCTGAC 713

     20 rThrLeuProAspArgLeuProProHisIleThrThrLeuValIleProA 37
        |::::::|||::   |||:::::   :::   :::|||   |||:::
    714 GGCCGTGCCCGAGGGGCTCAGCGCCTTCACCCAAGCGCTGGATATCAGTA 763

     37 spAsnAsnLeuThrSerLeuProGluLeuProGluGlyLeuArgGluLeu 53
          ||||||:::|||   |||||||||       :::::::::::::::::
    764 TGAACAACATTACTCAGTTGCCAGAA......GATGCATTTAAGAACTTT 807

     54 GluValSerGlyAsnLeuGlnLeuThrSer......LeuProSerLeu.. 67
                  :::||||||||||::::::      |||:::   :::
    808 CCTTTTCTAGAAGAGCTACAATTGGCGGGCAACGACCTTTCTTTTATCCA 857

     68 .ProGlnGlyLeuGlnLysLeuTrpAlaTyrAsnAsnTrpLeuAlaSerL 84
      |||::::::|||      |||
    858 CCCAAAGGCCTTGTCTGGGTTG........................... 879

     84 euProThrLeuProProGlyLeuGlyAspLeuAlaValSerAsnAsnGln 100
                        |||      |||::::::   |||||||||
    880 ..............AAAGAACTCAAAGTTCTAACGCTCCAGAATAATCAG 915

    101 LeuThrSerLeuProGluMetPro............ProAlaLeuArgGl 113
        |||::::::::::|||      :::           :::|||||||:::
    916 TTGAAAACAGTACCCAGTGAAGCCATTCGAGGGCTGAGTGCTTTGCAGTC 965

    113 uLeuArgValSerGlyAsnAsnLeuThrSerLeuArgAlaLeuPro.... 128
        ||||||:::::::::|||:::::|||||||:::   :::   :::
    966 TTTGCGTTTAGATGCCAACCATATTACCTCAGTCCCCGAGGACAGTTTTG 1015

    129 ........SerGlyLeuGlnLysLeuTrpAlaTyrAsnAsnArgLeuThr 142
                  |||:::   ||||||      :::|||   ||||||
   1016 AAGGACTTGTTCAGTTACGGCATCTGTGGCTGGATGACAACAGCTTGACG 1065

    143 SerLeuProGluMetSer............ProGlyLeuGlnGluLeuAs 155
          :::|||      :::            |||   ||||||:::|||
```

```
    1056 GAGGTGCCTGTGCACCCCCTCAGCAATCTGCCCACCCTACAGGCGCTGAC 1115

     155 pValSerHisAsnGlnLeuThrArgLeuProGln...SerLeuThrGlyL 171
         ::::::   |||:::::::::   :::|||:::   ::::::|||:::|
    1116 CCTGGCTCTCAACAAGATCTCAAGTATCCCTGACTTTGCATTTACCAACC 1165

     171 euSerSerAlaAlaArgValTyrLeuAspGlyAsnProLeuSerValArg 187
         ||||||||   :::   ::::::|||:::::|||         :::|||
    1166 TTTCAAGCCTGGTAGTTCTGCATCTTCATAACAAT......AAAATTAGA 1209

     188 ThrArgAspArgLeuCys 193
         :::   ::::::   |||
    1210 AGCCTGAGTCAACACTGT 1227

seq_name: N_Geneseq_36:X24014

seq_documentation_block:
ID   X24014 standard; cDNA; 3949 BP.
AC   X24014;
DT   29-JUN-1999  (first entry)
DE   Human AOMF05 cDNA.
KW   AOMF05; human; G-protein coupled glycoprotein hormone receptor; brain;
KW   anti-obesity; anti-diabetic; treatment; pancreatic disease; obesity;
KW   diabetes; endocrine system; skeletal muscle; spinal cord; placenta;
KW   transgenic animal; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             568. .3423
FT                   /*tag= a
FT                   /product= "AOMF05"
PN   WO9915545-A1.
PD   01-APR-1999.
PF   24-SEP-1998; U20101.
PR   24-SEP-1997; US-059868.
PA   (MERI ) MERCK & CO INC.
PI   Abramovitz M, Liu Q, McDonald TP, ONeill GP, Wang R;
DR   WPI; 99-254689/21.
DR   P-PSDB; W93906.
PT   Human G-protein coupled glycoprotein hormone receptor AOMF05
PS   Disclosure; Fig 6A-F; 89pp; English.
CC   This invention describes the isolation of a novel human G-protein coupled
CC   glycoprotein hormone receptor, AOMF05 which has anti-obesity and
CC   anti-diabetic activity. The AOMF05 gene, receptor and modulators are
CC   useful for treatment of pancreatic disease, obesity and diabetes.
CC   Glycoprotein hormone receptors are important in the endocrine system
CC   and AOMF05 may be involved in development and function of the skeletal
CC   muscle, spinal cord, placenta and to a lesser extent, the brain. The
CC   transgenic animal may be useful for studying tissue and temporal
CC   specific expression or activity of the AOMF05 receptor, as well as
CC   for studying the ability of a variety of compounds to act as modulators
CC   of AOMF05 receptor activity.
SQ   Sequence   3949 BP;    1016 A;    912 C;    950 G;    1071 T;


alignment_scores:
            Quality:  143.00              Length:    206
              Ratio:  1.144                 Gaps:      9
 Percent Similarity:  60.680   Percent Identity: 27.670

alignment_block:
 US-09-068-804-14 x X24014   ..

 Align seg 1/1  to: X24014  from: 1  to: 3949

       4 CysLeuAsnAsnGlyAsnProValLeuAsnValGlyAlaSerGlyLeuTh 20
         |||      :::|||:::     ::::::  ::::::  ||||||||
     664 TGCAGCTGCGACGGCGACCGTCGGGTGGACTGCTCCGGGAAGGGGCTGAC 713

      20 rThrLeuProAspArgLeuProProHisIleThrThrLeuValIleProA 37
         |:::::|||:::  |||::::::  :::  :::|||  |||:::
     714 GGCCGTGCCCGAGGGGCTCAGCGCCTTCACCCAAGCGCTGGATATCAGTA 763

      37 spAsnAsnLeuThrSerLeuProGluLeuProGluGlyLeuArgGluLeu 53
         ||||||:::|||  ||||||||      ::::::::::::::::::
     764 TGAACAACATTACTCAGTTGCCAGAA......GATGCATTTAAGAACTTT 807

      54 GluValSerGlyAsnLeuGlnLeuThrSer......LeuProSerLeu.. 67
                 :::|||||||||:::::      |||:::   :::
     808 CCTTTTCTAGAAGAGCTACAATTGGCGGGCAACGACCTTTCTTTTATCCA 857

      68 .ProGlnGlyLeuGlnLysLeuTrpAlaTyrAsnAsnTrpLeuAlaSerL 84
          |||::::::|||      |||
     858 CCCAAAGGCCTTGTCTGGGTTG.......................... 879

      84 euProThrLeuProProGlyLeuGlyAspLeuAlaValSerAsnAsnGln 100
                           |||      |||::::::   |||||||||
     880 .............AAAGAACTCAAAGTTCTAACGCTCCAGAATAATCAG 915

     101 LeuThrSerLeuProGluMetPro...........ProAlaLeuArgGl 113
         |||:::::::::|||      :::            :::||||||:::
     916 TTGAAAACAGTACCCAGTGAAGCCATTCGAGGGCTGAGTGCTTTGCAGTC 965

     113 uLeuArgValSerGlyAsnAsnLeuThrSerLeuArgAlaLeuPro.... 128
         ||||||:::::::::|||::::::|||||||:::   :::   :::
     966 TTTGCGTTTAGATGCCAACCATATTACCTCAGTCCCCGAGGACAGTTTTG 1015

     129 ........SerGlyLeuGlnLysLeuTrpAlaTyrAsnAsnArgLeuThr 142
                    |||:::   ||||||      ::::||   ||||||
    1016 AAGGACTTGTTCAGTTACGGCATCTGTGGCTGGATGACAACAGCTTGACG 1065

     143 SerLeuProGluMetSer...........ProGlyLeuGlnGluLeuAs 155
         :::|||      :::            |||   ||||||:::|||
    1066 GAGGTGCCTGTGCACCCCCTCAGCAATCTGCCCACCCTACAGGCGCTGAC 1115

     155 pValSerHisAsnGlnLeuThrArgLeuProGln...SerLeuThrGlyL 171
         ::::::   |||:::::::::   :::|||:::   ::::::|||:::|
    1116 CCTGGCTCTCAACAAGATCTCAAGTATCCCTGACTTTGCATTTACCAACC 1165

     171 euSerSerAlaAlaArgValTyrLeuAspGlyAsnProLeuSerValArg 187
         ||||||||   :::   ::::::|||:::::|||         :::|||
    1166 TTTCAAGCCTGGTAGTTCTGCATCTTCATAACAAT......AAAATTAGA 1209

     188 ThrArgAspArgLeuCys 193
         :::   ::::::   |||
    1210 AGCCTGAGTCAACACTGT 1227

seq_name: N_Geneseq_36:X24013

seq_documentation_block:
ID   X24013 standard; cDNA; 3950 BP.
AC   X24013;
DT   29-JUN-1999  (first entry)
DE   Human AOMF05 cDNA.
KW   AOMF05; human; G-protein coupled glycoprotein hormone receptor; brain;
KW   anti-obesity; anti-diabetic; treatment; pancreatic disease; obesity;
KW   diabetes; endocrine system; skeletal muscle; spinal cord; placenta;
KW   transgenic animal; ss.
OS   Homo sapiens.
PN   WO9915545-A1.
PD   01-APR-1999.
PF   24-SEP-1998; U20101.
PR   24-SEP-1997; US-059868.
PA   (MERI ) MERCK & CO INC.
PI   Abramovitz M, Liu Q, McDonald TP, ONeill GP, Wang R;
DR   WPI; 99-254689/21.
PT   Human G-protein coupled glycoprotein hormone receptor AOMF05
PS   Claim 4c; Fig 4A-B; 89pp; English.
CC   This invention describes the isolation of a novel human G-protein coupled
CC   glycoprotein hormone receptor, AOMF05 which has anti-obesity and
CC   anti-diabetic activity. The AOMF05 gene, receptor and modulators are
CC   useful for treatment of pancreatic disease, obesity and diabetes.
CC   Glycoprotein hormone receptors are important in the endocrine system
CC   and AOMF05 may be involved in development and function of the skeletal
CC   muscle, spinal cord, placenta and to a lesser extent, the brain. The
CC   transgenic animal may be useful for studying tissue and temporal
CC   specific expression or activity of the AOMF05 receptor, as well as
```

```
CC   for studying the ability of a variety of compounds to act as modulators
CC   of AOMF05 receptor activity.
SQ   Sequence   3950 BP;    1017 A;    912 C;    950 G;    1071 T;


alignment_scores:
            Quality:  143.00              Length:    206
              Ratio:  1.144                 Gaps:      9
 Percent Similarity:  60.680    Percent Identity: 27.670

alignment_block:
 US-09-068-804-14 x X24013   ..

 Align seg 1/1  to: X24013  from: 1  to: 3950

       4 CysLeuAsnAsnGlyAsnProValLeuAsnValGlyAlaSerGlyLeuTh 20
         |||      :::|||:::      ::::::   ::::::   ||||||||
     665 TGCAGCTGCGACGGCGACCGTCGGGTGGACTGCTCCGGGAAGGGGCTGAC 714

      20 rThrLeuProAspArgLeuProProHisIleThrThrLeuValIleProA 37
         |:::::|||:::   |||:::::   :::   :::|||   |||:::
     715 GGCCGTGCCCGAGGGGCTCAGCGCCTTCACCCAAGCGCTGGATATCAGTA 764

      37 spAsnAsnLeuThrSerLeuProGluLeuProGluGlyLeuArgGluLeu 53
           ||||||:::|||   ||||||||      :::::::::::::::::
     765 TGAACAACATTACTCAGTTGCCAGAA......GATGCATTTAAGAACTTT 808

      54 GluValSerGlyAsnLeuGlnLeuThrSer......LeuProSerLeu.. 67
                    :::|||||||||:::::      |||:::   :::
     809 CCTTTTCTAGAAGAGCTACAATTGGCGGGCAACGACCTTTCTTTTATCCA 858

      68 .ProGlnGlyLeuGlnLysLeuTrpAlaTyrAsnAsnTrpLeuAlaSerL 84
          |||:::::::|||      |||
     859 CCCAAAGGCCTTGTCTGGGTTG.......................... 880

      84 euProThrLeuProProGlyLeuGlyAspLeuAlaValSerAsnAsnGln 100
                          |||      |||::::::   ||||||||
     881 ..............AAAGAACTCAAAGTTCTAACGCTCCAGAATAATCAG 916

     101 LeuThrSerLeuProGluMetPro............ProAlaLeuArgGl 113
         |||:::::::::|||      :::           :::|||||:::
     917 TTGAAAACAGTACCCAGTGAAGCCATTCGAGGGCTGAGTGCTTTGCAGTC 966

     113 uLeuArgValSerGlyAsnAsnLeuThrSerLeuArgAlaLeuPro.... 128
          ||||||:::::::::|||:::::|||||:::   :::   :::
     967 TTTGCGTTTAGATGCCAACCATATTACCTCAGTCCCCGAGGACAGTTTTG 1016

     129 ........SerGlyLeuGlnLysLeuTrpAlaTyrAsnAsnArgLeuThr 142
                    |||:::   ||||||      :::|||   ||||||
    1017 AAGGACTTGTTCAGTTACGGCATCTGTGGCTGGATGACAACAGCTTGACG 1066

     143 SerLeuProGluMetSer............ProGlyLeuGlnGluLeuAs 155
         :::|||      :::            |||   |||||:::|||
    1067 GAGGTGCCTGTGCACCCCCTCAGCAATCTGCCCACCCTACAGGCGCTGAC 1116

     155 pValSerHisAsnGlnLeuThrArgLeuProGln...SerLeuThrGlyL 171
          ::::::   |||:::::::::   :::|||:::   ::::::|||:::|
    1117 CCTGGCTCTCAACAAGATCTCAAGTATCCCTGACTTTGCATTTACCAACC 1166

     171 euSerSerAlaAlaArgValTyrLeuAspGlyAsnProLeuSerValArg 187
         ||||||||   :::   ::::::|||:::::::|||         :::|||
    1167 TTTCAAGCCTGGTAGTTCTGCATCTTCATAACAAT......AAAATTAGA 1210

     188 ThrArgAspArgLeuCys 193
         :::   ::::::   |||
    1211 AGCCTGAGTCAACACTGT 1228

 seq_name: N_Geneseq_36:X24011

 seq_documentation_block:
 ID   X24011 standard; cDNA; 3993 BP.
 AC   X24011;
 DT   29-JUN-1999  (first entry)
 DE   Human AOMF05 cDNA.
 KW   AOMF05; human; G-protein coupled glycoprotein hormone receptor; brain;
 KW   anti-obesity; anti-diabetic; treatment; pancreatic disease; obesity;
 KW   diabetes; endocrine system; skeletal muscle; spinal cord; placenta;
 KW   transgenic animal; ss.
 OS   Homo sapiens.
 PN   WO9915545-A1.
 PD   01-APR-1999.
 PF   24-SEP-1998; U20101.
 PR   24-SEP-1997; US-059868.
 PA   (MERI ) MERCK & CO INC.
 PI   Abramovitz M, Liu Q, McDonald TP, ONeill GP, Wang R;
 DR   WPI; 99-254689/21.
 PT   Human G-protein coupled glycoprotein hormone receptor AOMF05
 PS   Claim 1; Fig 1A-B; 89pp; English.
 CC   This invention describes the isolation of a novel human G-protein coupled
 CC   glycoprotein hormone receptor, AOMF05 which has anti-obesity and
 CC   anti-diabetic activity. The AOMF05 gene, receptor and modulators are
 CC   useful for treatment of pancreatic disease, obesity and diabetes.
 CC   Glycoprotein hormone receptors are important in the endocrine system
 CC   and AOMF05 may be involved in development and function of the skeletal
 CC   muscle, spinal cord, placenta and to a lesser extent, the brain. The
 CC   transgenic animal may be useful for studying tissue and temporal
 CC   specific expression or activity of the AOMF05 receptor, as well as
 CC   for studying the ability of a variety of compounds to act as modulators
 CC   of AOMF05 receptor activity.
 SQ   Sequence   3993 BP;    1022 A;    923 C;    960 G;    1085 T;


alignment_scores:
            Quality:  143.00              Length:    206
              Ratio:  1.144                 Gaps:      9
 Percent Similarity:  60.680    Percent Identity: 27.670

alignment_block:
 US-09-068-804-14 x X24011   ..

 Align seg 1/1  to: X24011  from: 1  to: 3993

       4 CysLeuAsnAsnGlyAsnProValLeuAsnValGlyAlaSerGlyLeuTh 20
         |||      :::|||:::      ::::::   ::::::   ||||||||
     665 TGCAGCTGCGACGGCGACCGTCGGGTGGACTGCTCCGGGAAGGGGCTGAC 714

      20 rThrLeuProAspArgLeuProProHisIleThrThrLeuValIleProA 37
         |:::::|||:::   |||:::::   :::   :::|||   |||:::
     715 GGCCGTGCCCGAGGGGCTCAGCGCCTTCACCCAAGCGCTGGATATCAGTA 764

      37 spAsnAsnLeuThrSerLeuProGluLeuProGluGlyLeuArgGluLeu 53
           ||||||:::|||   ||||||||      :::::::::::::::::
     765 TGAACAACATTACTCAGTTGCCAGAA......GATGCATTTAAGAACTTT 808

      54 GluValSerGlyAsnLeuGlnLeuThrSer......LeuProSerLeu.. 67
                    :::|||||||||:::::      |||:::   :::
     809 CCTTTTCTAGAAGAGCTACAATTGGCGGGCAACGACCTTTCTTTTATCCA 858

      68 .ProGlnGlyLeuGlnLysLeuTrpAlaTyrAsnAsnTrpLeuAlaSerL 84
          |||:::::::|||      |||
     859 CCCAAAGGCCTTGTCTGGGTTG.......................... 880

      84 euProThrLeuProProGlyLeuGlyAspLeuAlaValSerAsnAsnGln 100
                          |||      |||::::::   ||||||||
     881 ..............AAAGAACTCAAAGTTCTAACGCTCCAGAATAATCAG 916

     101 LeuThrSerLeuProGluMetPro............ProAlaLeuArgGl 113
         |||:::::::::|||      :::           :::|||||:::
     917 TTGAAAACAGTACCCAGTGAAGCCATTCGAGGGCTGAGTGCTTTGCAGTC 966

     113 uLeuArgValSerGlyAsnAsnLeuThrSerLeuArgAlaLeuPro.... 128
          ||||||:::::::::|||:::::|||||:::   :::   :::
     967 TTTGCGTTTAGATGCCAACCATATTACCTCAGTCCCCGAGGACAGTTTTG 1016

     129 ........SerGlyLeuGlnLysLeuTrpAlaTyrAsnAsnArgLeuThr 142
                    |||:::   ||||||      :::|||   ||||||
```

```
     1017 AAGGACTTGTTCAGTTACGGCATCTGTGGCTGGATGACAACAGCTTGACG 1066

      143 SerLeuProGluMetSer.....ProGlyLeuGlnGluLeuAs 155
          :::|||      :::          |||   ||||||:::|||
     1067 GAGGTGCCTGTGCACCCCCTCAGCAATCTGCCCACCCTACAGGCGCTGAC 1116

      155 pValSerHisAsnGlnLeuThrArgLeuProGln...SerLeuThrGlyL 171
          ::::::   |||:::::::::   :::|||:::   ::::::|||:::|
     1117 CCTGGCTCTCAACAAGATCTCAAGTATCCCTGACTTTGCATTTACCAACC 1166

      171 euSerSerAlaAlaArgValTyrLeuAspGlyAsnProLeuSerValArg 187
          ||||||||   :::   ::::::::|||:::::::|||         :::|||
     1167 TTTCAAGCCTGGTAGTTCTGCATCTTCATAACAAT......AAAATTAGA 1210

      188 ThrArgAspArgLeuCys 193
          :::   ::::::   |||
     1211 AGCCTGAGTCAACACTGT 1228
```

seq_name: N_Geneseq_36:T16517

seq_documentation_block:
ID T16517 standard; DNA; 4192 BP.
AC T16517;
DT 01-OCT-1996 (first entry)
DE Collagen A1/decorin chimeric gene.
KW Decorin; dermatan sulfate proteoglycan; collagen IA; scar;
KW tissue repair; ds.
OS Synthetic.
FH Key Location/Qualifiers
FT cds 20. .4186
FT /*tag= a
FT /product= collagen IA/decorin fusion protein
FT misc_difference 2679
FT /*tag= b
FT /note= "base 2679 is not identified in the
FT specification"
FT misc_difference 2688
FT /*tag= c
FT /note= "base 2688 is not identified in the
FT specification"
PN CA2151547-A.
PD 11-DEC-1995.
PF 12-JUN-1995; 151547.
PR 10-JUN-1994; US-259263.
PA (USSU) US SURGICAL CORP.
PI Espino P, Gruskin EA;
DR WPI; 96-140144/15.
DR P-PSDB; R89471.
PT Chimaeric DNA encoding protein contg. extracellular matrix protein
PT domain - and cellular regulatory factor domain, partic. useful as
PT osteogenic agents, also related vectors, transformed cells and
PT chimaeric proteins.
PS Disclosure; Fig 3; 59pp; English.
CC A chimeric gene (T16517) codes for a fusion protein (R89471) between
CC the helical region of human collagen I(a) and human dermatan
CC sulphate proteoglycan (decorin). The collagen moiety was cloned
CC from human fibroblast AG02261A cells by PCR amplification. The
CC construct was inserted into a pMal vector for expression in E. coli.
CC The fusion protein provides sustained release and delivery of
CC decorin to a target tissue. The decorin acts to reduce scarring
CC of healing tissue.
SQ Sequence 4192 BP; 722 A; 1275 C; 1306 G; 887 T;

alignment_scores:
Quality: 139.50 Length: 260
Ratio: 0.962 Gaps: 14
Percent Similarity: 55.769 Percent Identity: 25.769

alignment_block:
US-09-068-804-14 x T16517 ..

Align seg 1/1 to: T16517 from: 1 to: 4192

```
       11 ValLeuAsnValGlyAlaSerGlyLeuThrThrLeuProAspArgLeuPr 27
          |||::::::   ::::::   ||||||   ::::::|||      |||||
     3296 GTGGTCCAGTGTTCTGATTTGGGTCTGGACAAAGTGCCAAAGGATCTTCC 3345

       27 oProHisIleThrThrLeuValIleProAspAsnAsnLeuThrSerLeuP 44
          ||||:::::|||   |||   :::::::::|||:::::|||
     3346 CCCTGACACAACTCTGCTAGACCTGCAAAACAACAAAATAACC..... 3388

       44 roGluLeuProGluGly.....LeuArgGluLeuGlu..... 54
          |||:::   :::|||          |||:::::|||:::
     3389 ..GAAATCAAAGATGGAGACTTTAAGAACCTGAAGAACCTTCACGCATTG 3436

       55 ......ValSerGlyAsnLeu..... 59
                |||:::::::::::
     3437 ATTCTTGTCAACAATAAAATTAGCAAAGTTAGTCCTGGAGCATTTACACC 3486

       60 .....GlnLeuThrSerL 64
                          |||||||:::   |
     3487 TTTGGTGAAGTTGGAACGACTTTATCTGTCCAAGAATCAGCTGAAGGAAT 3536

       64 euPro...SerLeuProGlnGlyLeuGlnLysLeuTrpAlaTyrAsnAsn 79
          |||||      :::||||:::   ||||||   |||:::|||:::::|||
     3537 TGCCAGAAAAAATGCCCAAAACTCTTCAGGAGCTGCGTGCCCATGAGAAT 3586

       80 TrpLeuAlaSerLeuProThrLeuPro...ProGlyLeuGlyAspLeuAl 95
          ::::::   :::   :::::::::      ||||||:::::::::
     3587 GAGATCACCAAAGTGCGAAAAGTTACTTTCAATGGACTGAACCAGATGAT 3636

       95 aVal..... 96
           |||
     3637 TGTCATAGAACTGGGCACCAATCCGCTGAAGAGCTCAGGAATTGAAAATG 3686

       97 .....SerAsnAsn 99
                        :::::::::
     3687 GGGCTTTCCAGGGAATGAAGAAGCTCTCCTACATCCGCATTGCTGATACC 3736

      100 GlnLeuThrSerLeuProGlu...MetProProAlaLeuArgGluLeuAr 115
          ::::::||||||:::|||:::   :::|||||||:::|||   ||||||::
     3737 AATATCACCAGCATTCCTCAAGGTCTTCCTCCTTCCCTTACGGAATTACA 3786

      115 gValSerGlyAsnAsnLeu.....ThrSerLeuArgAlaL 127
          :::::::||||||:::::          :::||||||:::::|
     3787 TCTTGATGGCAACAAAATCAGCAGAGTTGATGCAGCTAGCCTGAAAGGAC 3836

      127 euProSerGlyLeuGlnLysLeu...TrpAlaTyrAsnAsnArgLeuThr 142
          ||   :::::|||   ||||||      ::::::|||:::      :::
     3837 TG...AATAATTTGGCTAAGTTGGGATTGAGTTTCAACAGCATCTCTGCT 3883

      143 .....SerLeuProGluMetSerProGlyLeuGlnGluLeuAs 155
                |||||||:::::   :::|||   |||:::|||||::
     3884 GTTGACAATGGCTCTCTGGCCAAC...ACGCCTCATCTGAGGGAGCTTCA 3930

      155 pValSerHisAsnGlnLeuThrArgLeuProGlnSerLeuThrGlyLeuS 172
          ::::::::::|||:::|||||||||:::|||   :::|||:::
     3931 CTTGGACAACAACAAGCTTACCAGAGTACCTGGTGGGCTGGCAGAGCATA 3980

      172 erSerAlaAlaArgValTyrLeuAspGlyAsnProLeuSerValArgThr 188
                |||||||||:::::::|||   :::||||||
     3981 AGTACATCCAGGTTGTCTACCTTCATAACAACAATATCTCTGTAGTTGGA 4030

      189 ArgAspArgLeuCys.....GlyHisHis 196
          :::   :::|||      ||||||:::
     4031 TCAAGTGACTTCTGCCCACCTGGACACAAC 4060
```

seq_name: N_Geneseq_36:Q50051

seq_documentation_block:
ID Q50051 standard; DNA; 711 BP.
AC Q50051;
DT 28-APR-1994 (first entry)
DE Decorin (N-terminal to LRR8) coding sequence PT-76.
KW leucine-rich repeat; proteoglycan; cell regulatory factor; MBP;

```
KW   fusion protein; maltose binding protein; tumour growth; inhibition;
KW   decorin; PG-II; PG-40; transforming growth factor-beta; TGF-beta; ds.
FH   Key             Location/Qualifiers
FT   cds             1. .705
FT                   /*tag= a
FT                   /product= PT-76
FT                   /note= "encodes decorin N-terminal sequence
FT                   and first 8 leucine-rich regions"
PN   WO9320202-A.
PD   14-OCT-1993.
PF   02-APR-1993; U03171.
PR   03-APR-1992; US-865652.
PA   (LJOL-) LA JOLLA CANCER RES FOUND.
PI   Cardenas J,  Craig W,  Mullen DG,  Pierschbacher MD;
PI   Ruoslahti EI;
DR   WPI; 93-336910/42.
DR   P-PSDB; R42265.
PT   Active fragments of protein esp. decorin - with cell regulatory
PT   factor domain, useful for inhibiting cell regulatory factor
PT   activity
PS   Example 5; Page 45-46; 77pp; English.
CC   Active fragments of decorin (full-length coding sequence Q50046)
CC   were generated by PCR and fused to Maltose Binding Protein. The
CC   resulting fusion proteins were useful for inhibiting the activity of
CC   a cell regulatory factor, esp. TGF-beta, and hence for treating
CC   conditions associated with over-activity of the growth factor such
CC   as certain tumours.
SQ   Sequence   711 BP;    213 A;    167 C;    155 G;    176 T;


alignment_scores:
            Quality:  127.00            Length:    216
              Ratio:  1.050               Gaps:      9
 Percent Similarity:  56.019   Percent Identity: 25.463

alignment_block:
 US-09-068-804-14 x Q50051   ..

 Align seg 1/1  to: Q50051  from: 1  to: 711

      11 ValLeuAsnValGlyAlaSerGlyLeuThrThrLeuProAspArgLeuPr 27
         |||::::::   ::::::   ||||||   :::::::||      |||||
     106 GTGGTCCAGTGTTCTGATTTGGGTCTGGACAAAGTGCCAAAGGATCTTCC 155

      27 oProHisIleThrThrLeuValIleProAspAsnAsnLeuThrSerLeuP 44
         ||||:::::::|||   |||   :::::::::|||:::::::|||   :::
     156 CCCTGACACAACTCTGCTAGACCTGCAAAACAACAAAATAACCGAAATCA 205

      44 roGluLeuProGluGlyLeuArgGluLeuGluValSerGlyAsnLeuGln 60
         :::         :::::::::::::|||           :::|||
     206 AAGAT......GGAGACTTTAAGAACCTGAAGAACCTTCACGCATTGATT 249

      61 LeuThrSerLeuProSerLeuProGlnGlyLeuGlnLysLeuTrpAlaTy 77
         |||::::::
     250 CTTGTCAAC..................................... 258

      77 rAsnAsnTrpLeuAlaSerLeuProThrLeuProProGly..... 90
                    :::   ::::::::::::::::||||||
     259 .............AATAAAATTAGCAAAGTTAGTCCTGGAGCATTTACAC 295

      91 ...........LeuGlyAspLeuAlaValSerAsnAsnGlnLeuThrSer 103
                    |||       |||   :::|||:::||||||||||:::
     296 CTTTGGTGAAGTTGGAACGACTTTATCTGTCCAAGAATCAGCTGAAGGAA 345

     104 LeuProGlu...MetProProAlaLeuArgGluLeuArgValSerGlyAs 119
         |||||||||   ||||||   :::|||:::||||||||||:::     ||
     346 TTGCCAGAAAAAATGCCCAAAACTCTTCAGGAGCTGCGTGCCCATGAGAA 395

     119 nAsnLeuThrSerLeuArgAlaLeuPro..................... 128
         |:::::|||   :::|||   ::::::
     396 TGAGATCACCAAAGTGCGAAAAGTTACTTTCAATGGACTGAACCAGATGA 445

     128 ..................................................... 128

     446 TTGTCATAGAACTGGGCACCAATCCGCTGAAGAGCTCAGGAATTGAAAAT 495

     129 .........SerGlyLeuGlnLysLeuTrpAlaTyr.............As 138
                    |||::::::||||||   :::|||             ::
     496 GGGGCTTTCCAGGGAATGAAGAAGCTC...TCCTACATCCGCATTGCTGA 542

     138 nAsnArgLeuThrSerLeuProGlu...MetSerProGlyLeuGlnGluL 154
         ::::   :::||||||:::|||:::   ::::::|||:::|||   ||||
     543 TACCAATATCACCAGCATTCCTCAAGGTCTTCCTCCTTCCCTTACGGAAT 592

     154 euAspValSerHisAsnGlnLeuThrArgLeu...ProGlnSerLeuThr 169
         ||::::::::::   |||::::::::::|||:::   :::   ||||||:::
     593 TACATCTTGATGGCAACAAAATCAGCAGAGTTGATGCAGCTAGCCTGAAA 642

     170 GlyLeuSerSerAlaAlaArgValTyrLeuAspGlyAsnProLeuSer 185
         ||||||::::::   |||:::::::   |||:::   |||::::::|||
     643 GGACTGAATAATTTGGCTAAGTTGGGATTGAGTTTCAACAGCATCTCT 690


seq_name: N_Geneseq_36:T08768

seq_documentation_block:
ID   T08768 standard; cDNA; 2446 BP.
AC   T08768;
DT   20-MAR-1996  (first entry)
DE   Rat biglycan cDNA.
KW   Biglycan; proteoglycan; chondroitin sulphate; neuron protection;
KW   neurotrophic; central nervous system; CNS; memory loss; dementia;
KW   learning; ss.
OS   Rattus sp.
FH   Key             Location/Qualifiers
FT   cds             123. .1232
FT                   /*tag= a
FT   signal_peptide  123. .233
FT                   /*tag= b
FT   mat_peptide     234. .1229
FT                   /*tag= c
PN   WO9530432-A1.
PD   16-NOV-1995.
PF   09-MAY-1994; E01479.
PR   09-MAY-1994; WO-E01479.
PA   (BOEF ) BOEHRINGER MANNHEIM GMBH.
PI   Hasenoehrl R,  Huston J,  Junghans U,  Kappler J,  Koops A;
PI   Mueller HW;
DR   WPI; 95-403938/51.
DR   P-PSDB; R87951.
PT   Proteoglycan cpds., partic. chondroitin sulphate proteoglycan(s)
PT   for maintain structural and function of the CNS and attenuating
PT   memory deficit(s) in the elderly and patients with dementia
PS   Disclosure; Page 44-45; 60pp; English.
CC   A cDNA sequence (T08768) obtd. from rat smooth muscle codes for rat
CC   biglycan (R87951), a chondroitin sulphate proteoglycan with
CC   neurotrophic activity for brain neurons.  The cDNA may be inserted
CC   into a vector and used for prodn. of recombinant biglycan in
CC   eukaryotic host cells.
SQ   Sequence   2446 BP;    562 A;    768 C;    508 G;    608 T;


alignment_scores:
            Quality:  128.00            Length:    254
              Ratio:  1.008               Gaps:      9
 Percent Similarity:  50.000   Percent Identity: 22.047

alignment_block:
 US-09-068-804-14 x T08768   ..

 Align seg 1/1  to: T08768  from: 1  to: 2446

      11 ValLeuAsnValGlyAlaSerGlyLeuThrThrLeuProAspArgLeuPr 27
         |||::::::   ::::::   ||||||:::|||:::|||      :::::
     342 GTTGTTCAGTGCTCTGACTTGGGTCTGAAGACTGTGCCCAAGGAGATCTC 391

      27 oProHisIleThrThrLeuValIleProAspAsnAsnLeuThrSerLeu. 43
```

```
     :|||::::::|||   |||   :::::::::|||:::::::::   |||
 392 ACCTGATACCACACTGCTAGACCTGCAGAACAATGACATCTCTGAACTCC 441

  43 ..................................................  43

 442 GCAAGGATGACTTCAAAGGCCTCCAGCACCTCTATGCTCTGGTCCTGGTA 491

  44 ............ProGluLeuProGlu............GlyLeuArgGl  52
                  :::   :::   |||              ||||||
 492 AACAATAAGATCTCCAAGATCCATGAGAAGGCCTTTAGCCCTCTGCGGAA 541

  52 uLeuGlu.........ValSerGlyAsnLeuGlnLeuThrSerLeuProS  66
      |||:::         :::|||   |||      :::          |||:
 542 GCTGCAGAAACTCTACATCTCCAAGAACCACCTGGTGGAGATTCCTCCCA 591

  66 erLeuProGlnGlyLeuGlnLysLeuTrpAlaTyrAsnAsnTrpLeuAla  82
     ::||||||   :::|||      |||:::   :::::::|||::::::
 592 ACCTGCCCAGCTCCCTGGTAGAGCTACGAATCCATGACAACCGTATCCGC 641

  83 SerLeuProThr.....................................  86
        :::|||:::
 642 AAAGTGCCCAAGGGCGTGTTCAGTGGGCTCCGGAACATGAACTGCATTGA 691

  87 ...............................LeuProProGly.......  90
                                     :::   ||||||
 692 GATGGGTGGGAATCCCCTGGAGAACAGTGGCTTTGAACCCGGAGCCTTTG 741

  91 ...........LeuGlyAspLeuAlaValSerAsnAsnGlnLeuThrSer 103
                |||:::   |||   :::|||:::::::::|||||||:::
 742 ATGGCCTGAAGCTCAACTACCTGCGCATCTCAGAGGCCAAGCTCACTGGC 791

 104 LeuPro...GluMetProProAlaLeuArgGluLeuArgValSerGlyAs 119
     :::|||   ::::::|||   :::|||   ||||||:::::::::   ||
 792 ATCCCCAAAGATCTCCCTGAGACCCTGAATGAACTTCACTTGGACCACAA 841

 119 nAsnLeuThrSerLeuArgAlaLeuProSerGlyLeuGlnLysLeuTrpA 136
     |:::::::   ::::::                 |||:::   |||
 842 CAAAATCCAGGCTATC..............GAGTTGGAGGATCTACTTC 876

 136 laTyrAsnAsn..............................ArgLeu 141
       |||::::::                              |||:::
 877 GCTACTCCAAGCTGTACAGGCTGGGCTTAGGACACAATCAGATCCGAATG 926

 142 ThrSerLeuProGluMetSer......ProGlyLeuGlnGluLeuAspVa 156
     :::             :::|||      |||   |||:::||||||:::::
 927 ATTGAGAATGGGAGCCTGAGTTTTCTGCCTACCCTGAGGGAACTTCACTT 976

 156 lSerHisAsnGlnLeuThrArgLeuProGlnSerLeuThrGlyLeuSerS 173
     ::::::::|||:::|||:::|||:::|||   :::|||::::::|||
 977 GGACAACAACAAGCTGTCCCGGGTGCCTGCTGGCCTCCCAGATCTCAAAC 1026

 173 erAlaAlaArgValTyrLeuAspGlyAsnProLeuSerValArgThrArg 189
               |||||||||:::::::|||   ::::::
1027 TCCTCCAGGTTGTCTATCTGCACTCCAACAACATCACCAAGGTGGGCATC 1076

 190 AspArgLeuCys 193
     :::   :::|||
1077 AACGATTTCTGC 1088
```

This Page Blank (uspto)

```
  MEDLINE   90202708
COMMENT     Draft entry and computer-readable sequence [1] kindly submitted by
            A.B.Hartman, 12-FEB-1990.
FEATURES             Location/Qualifiers
     source          1. .2900
                     /organism="Shigella flexneri"
                     /db_xref="taxon:623"
     -35_signal      139. .144
     -10_signal      162. .167
     CDS             251. .1849
                     /note="invasion plasmid antigen H"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA26526.1"
                     /db_xref="PID:g152747"
                     /db_xref="GI:152747"
                     /translation="MGKELFSREERGIAFNRLSQCFQNQEAVLNLSDLNLTSLPELPK
                     HISALIVENNKLTSLPKLPAFLKELNADNNRLSVIPELPESLTTLSVRSNQLENLPVL
                     PNHLTSLFVENNRLYNLPALPEKLKFLHVYYNRLTTLPDLPDKLEILCAQRNNLVTFP
                     QFSDRNNIRQKEYYFHFNQITTLPESFSQLDSSYRINISGNPLSTRVLQSLQRLTSSP
                     DYHGPQIYFSMSDGQQNTLHRPLADAVTAWFPENKQSDVSQIWHAFEHEEHANTFSAF
                     LDRLSDTVSARNTSGFREQVAAWLEKLSASAELRQQSFAVAADATESCEDRVALTWNN
                     LRKTLLVHQASEGLFDNDTGALLSLGREMFRLEILEDIARDKVRTLHFVDEIEVYLAF
                     QTMLAEKLQLSTAVKEMRFYGVSGVTANDLRTAEAMVRSREENEFTDWFSLWGPWHAV
                     LKRTEADRWAQAEEQKYEMLENEYSQRVADRLKASGLSGDADAQREAGAQVMRETEQQ
                     IYRQLTDEVLALRLSENGSRLHHS"
     -35_signal      2111. .2116
     -10_signal      2132. .2137
     CDS             2277. .>2900
                     /note="ORF3"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA26527.1"
                     /db_xref="PID:g152748"
                     /db_xref="GI:152748"
                     /translation="MKPINNHSFFRSLCGLSCISRLSVEEQCTRDYHRIWDDWAREGT
                     TTENRIQAVRLLKICLDTREPVLNLSLLKLRSLPPLPLHIRELNISNNELISLPENSP
                     LLTELHVNGNNLNILPTLPSQLIKLNISFNRNLSCLPSLPPYLQSLSARFNSLETLPE
                     LPSTLTILRIEGNRLTVLPELPHRLQELFVSGNRLQELPEFPQSLKYL"
BASE COUNT      868 a    657 c    578 g    797 t
ORIGIN


alignment_scores:
            Quality:  853.00              Length:    533
              Ratio:  2.257                 Gaps:     10
 Percent Similarity:  70.919   Percent Identity: 38.462

alignment_block:
 US-09-068-804-14 x SHFIPAH   ..

 Align seg 1/1  to: SHFIPAH  from: 1  to: 2900

       1 MetArgAspCysLeuAsnAsnGlyAsnProValLeuAsnValGlyAlaSe 17
         :::   :::|||:::::::|||   ::::::|||||||||:::::::::
     302 CTCTCCCAGTGCTTTCAGAATCAAGAAGCAGTATTAAATTTATCAGACCT 351

      17 rGlyLeuThrThrLeuProAspArgLeuProProHisIleThrThrLeuV 34
         :::|||||||:::|||||||:::   ||||||   ||||||:::::::|||:
     352 AAATTTGACGTCTCTTCCCGAA...TTACCAAAGCATATTTCTGCTTTGA 398

      34 alIleProAspAsnAsnLeuThrSerLeuProGluLeuProGluGlyLeu 50
         :::::   :::|||:::|||||||||||||||||   |||||||:::   |||
     399 TTGTAGAAAATAATAAATTAACATCATTGCCAAAGCTGCCTGCATTTCTT 448

      51 ArgGluLeuGluValSerGlyAsnLeuGlnLeuThrSerLeuProSerLe 67
         :::|||||||:::::::::::::|||   :::|||:::   :::|||   ||
     449 AAAGAACTTAATGCTGATAATAAC...AGGCTTTCTGTGATACCAGAACT 495

      67 uProGlnGlyLeuGlnLysLeuTrpAlaTyrAsnAsnTrpLeuAlaSerL 84
         ||||:::::|||   :::|||   :::   :::|||   |||:::::::|
     496 TCCTGAGTCATTAACAACTTTAAGTGTTCGTTCTAATCAACTGGAAAACC 545

      84 euProThrLeuProProGlyLeuGlyAspLeuAlaValSerAsnAsnGln 100
         |||||:::||||||      |||   :::|||   |||   |||||||:::
     546 TTCCTGTTTTGCCAAACCATTTAACATCATTATTTGTTGAAAATAACAGG 595

     101 LeuThrSerLeuProGluMetProProAlaLeuArgGluLeuArgValSe 117
         |||   :::|||||||:::::::|||      |||:::   |||:::|||
     596 CTATATAACTTACCGGCTCTTCCCGAAAAATTGAAATTTTTACATGTTTA 645

     117 rGlyAsnAsnLeuThrSerLeuArgAlaLeuProSerGlyLeuGlnLysL 134
           |||   ||||||:::|||   :::|||||||:::   |||:::   |
     646 TTATAACAGGCTGACAACATTACCCGACTTACCGGATAAACTGGAAATTC 695

     134 euTrpAlaTyrAsnAsnArgLeuThrSerLeuProGluMetSerProGly 150
         ||   |||      |||   |||:::::::::::|||:::   |||
     696 TCTGTGCTCAGCGCAATAATCTGGTTACTTTTCCTCAATTTTCT..... 739

     151 LeuGlnGluLeuAspValSerHis.....................AsnGl 160
            :::      ::::::   :::                      |||||
     740 ...GATAGAAACAATATCAGACAAAAGGAATATTATTTTCATTTTAATCA 786

     160 nLeuThrArgLeuProGlnSerLeuThrGlyLeuSerSerAlaAlaArgV 177
         |:::|||   ||||||:::|||:::::::   |||:::|||:::   |||:
     787 GATAACCACTCTTCCGGAGAGTTTTTCACAATTAGATTCAAGTTACAGGA 836

     177 alTyrLeuAspGlyAsnProLeuSerValArgThrArgAspArgLeu... 192
         ::   :::::|||||||||||||||||:::|||:::   :::   |||
     837 TTAATATTTCAGGGAATCCATTGTCGACTCGCGTTCTGCAATCCCTGCAA 886

     193 .......................CysGlyHisHisTrpProPheArg.. 200
                                 |||   :::   |||
     887 AGATTAACCTCTTCGCCGGACTACCACGGCCCGCAGATTTACTTCTCCAT 936

     201 ..........HisGlnAspThrLeuArgTyrGlyGlyAlaPheArgProA 214
                 :::|||:::|||||||:::
     937 GAGTGACGGACAACAGAATACACTCCAT..................... 964

     214 rgGluAlaArgAlaLeuHisLeuAlaValAlaAspTrpLeuThrSerAla 230
               |||:::|||      ||||||:::::::|||:::::   :::
     965 ........CGCCCCCTGGCTGATGCCGTGACAGCATGGTTCCCGGAAAAC 1006

     231 ArgGluGlyGluAlaAlaGlnAlaAspArgTrpGlnAlaPheGlyLeuGl 247
         ::::::::::::::::::|||         |||:::|||||||      ||
    1007 AAACAATCTGATGTATCACAGATA......TGGCATGCTTTTGAACATGA 1050

     247 uAspAsnAlaAlaAlaPheSerLeuValLeuAspArgLeuArgGluThrG 264
         |:::::|||:::::|||||||      ||||||||||||   :::|||
    1051 AGAGCATGCCAACACCTTTTCCGCGTTCCTTGACCGCCTTTCCGATACCG 1100

     264 luAsnPheLysLysAspAlaGlyPheLysAlaGlnIleSerSerTrpLeu 280
           :::   ::::::   :::||||||:::::::|||:::::::::|||||
    1101 TCTCTGCACGCAATACCTCCGGATTCCGTGAACAGGTCGCTGCATGGCTG 1150

     281 ThrGlnLeuAlaGluAspAlaAlaLeuArgAlaLysThrPheAlaMetAl 297
         :::|||:::::::::|||:::||||||   ::::::||||||:::||
    1151 GAAAAACTCAGTGCCTCTGCGGAGCTTCGACAGCAGTCTTTCGCTGTTGC 1200

     297 aThrGluAlaThrSerThrCysGluAspArgValThrHisAlaLeuHisG 314
         |:::::||||||   :::|||||||||||||||||:::   :::   ::::
    1201 TGCTGATGCCACTGAGAGCTGTGAGGACCGTGTCGCGCTCACATGGAACA 1250

     314 lnMetAsnAsnValGlnLeuValHisAsnAlaGluLysGlyGluTyrAsp 330
         :::::   ::::::   |||||||||:::|||      |||   :::|||
    1251 ATCTCCGGAAAACCCTCCTGGTCCATCAGGCATCAGAAGGCCTTTTCGAT 1300

     331 AsnAsnLeuGlnGlyLeuValSerThrGlyArgGluMetPheArgLeuAl 347
         |||:::      :::|||:::|||   ||||||||||||||||||||::
    1301 AATGATACCGGCGCTCTGCTCTCCCTGGGCAGGGAAATGTTCCGCCTCGA 1350

     347 aThrLeuGluGlnIleAlaArgGluLysAlaGlyThrLeuAlaLeuValA 364
         :::|||||||:::|||||||||:::|||:::   ||||||   :::||||
    1351 AATTCTGGAGGACATTGCCCGGGATAAAGTCAGAACTCTCCATTTTGTGG 1400

     364 spAspValGluValTyrLeuAlaPheGlnAsnLysLeuLysGluSerLeu 380
```



seq_name: gb_ba2:AF047365

seq_documentation_block:

LOCUS AF047365 1822 bp DNA BCT 19-JUN-1998

DEFINITION Shigella flexneri plasmid pWR100 IpaH9.8 (ipaH9.8) gene, complete cds.

ACCESSION AF047365

NID g3237376

VERSION AF047365.1 GI:3237376

KEYWORDS .

SOURCE Shigella flexneri.

ORGANISM Shigella flexneri
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Shigella.

REFERENCE 1 (bases 1 to 1822)

AUTHORS Demers,B., Sansonetti,P.J. and Parsot,C.

TITLE Induction of type III secretion in shigella flexneri is associated with differential control of transcription of genes encoding secreted proteins

JOURNAL EMBO J. 17 (10), 2894-2903 (1998)

MEDLINE 98250661

REFERENCE 2 (bases 1 to 1822)

AUTHORS Demers,B., Sansonetti,P.J. and Parsot,C.

TITLE Direct Submission

JOURNAL Submitted (09-FEB-1998) Molecular Pathogenic Microbiology, Pasteur Institute, 28 rue du Dr. Roux, Paris 75724, France

FEATURES Location/Qualifiers

source 1. .1822
/organism="Shigella flexneri"
/strain="M90T"
/db_xref="taxon:623"
/plasmid="pWR100"
/note="virulence plasmid"

gene 160. .1797
/gene="ipaH9.8"

CDS 160. .1797
/gene="ipaH9.8"
/note="secreted by the type III secretion pathway encoded by the virulence plasmid"
/codon_start=1
/transl_table=11
/product="IpaH9.8"
/protein_id="AAC23714.1"
/db_xref="PID:g3237377"
/db_xref="GI:3237377"
/translation="MLPINNNFSLPQNSFYNTISGTYADYFSAWDKWEKQALPGEERD
EAVSRLKECLINNSDELRLDRLNLSSLPDNLPAQITLLNVSYNQLTNLPELPVTLKKL
YSASNKLSELPVLPPALESLQVQHNELENLPALPDSLLTMNISYNEIVSLPSLPQALK
NLRATRNFLTELPAFSEGNNPVVREYFFDRNQISHIPESILNLRNECSIHISDNPLSS
HALQALQRLTSSPDYHGPRIYFSMSDGQQNTLHRPLADAVTAWFPENKQSDVSQIWHA
FEHEEHANTFSAFLDRLSDTVSARNTSGFREQVAAWLEKLSASAELRQQSFAVAADAT
ESCEDRVALTWNNLRKTLLVHQASEGLFDNDTGALLSLGREMFRLEILEDIARDKVRT
LHFVDEIEVYLAFQTMLAEKLQLSTAVKEMRFYGVSGVTANDLRTAEAMVRSREENEF
TDWFSLWGPWHAVLKRTEADRWAQAEEQKYEMLENEYPQRVADRLKASGLSGDADAER
EAGAQVMRETEQQIYRQLTDEVLALRLSENGSQLHHS"

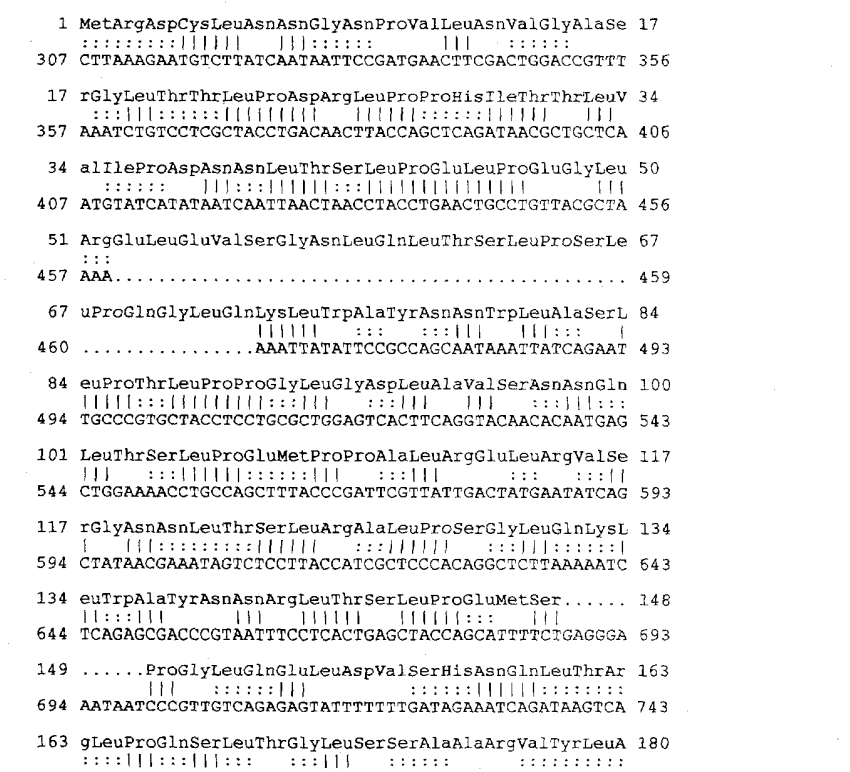
BASE COUNT 501 a 457 c 396 g 468 t

ORIGIN

alignment_scores:
Quality: 821.50 Length: 530
Ratio: 2.251 Gaps: 10
Percent Similarity: 68.868 Percent Identity: 36.981

alignment_block:
US-09-068-804-14 x AF047365 ..

Align seg 1/1 to: AF047365 from: 1 to: 1822



```
 744 TATCCCGGAAAGCATTCTTAATCTGAGGAATGAATGTTCAATACATATTA 793

 180 spGlyAsnProLeuSer.....ValArgThrArgAspArgLeuCys 193
     :::::|||||||||||||      :::::::::   :::|||||||
 794 GTGATAACCCATTATCATCCCATGCTCTGCAAGCCCTGCAAAGATTAACC 843

 194 Gly.....HisHisTrpPro.....PheArg..... 200
     :::        :::|||   |||        |||
 844 TCTTCGCCGGACTACCACGGCCCACGGATTTACTTCTCCATGAGTGACGG 893

 201 .HisGlnAspThrLeuArgTyrGlyGlyAlaPheArgProArgGluAlaA 217
      :::|||:::|||||||:::                              |
 894 ACAACAGAATACACTCCAT.............................C 913

 217 rgAlaLeuHisLeuAlaValAlaAspTrpLeuThrSerAlaArgGluGly 233
     ||:::|||     ||||||::::::|||:::::   ::::::::::::
 914 GCCCCCTGGCTGATGCCGTGACAGCATGGTTCCCGGAAAACAAACAATCT 963

 234 GluAlaAlaGlnAlaAspArgTrpGlnAlaPheGlyLeuGluAspAsnAl 250
     ::::::::::|||         |||:::||||||     |||:::::::||
 964 GATGTATCACAGATA......TGGCATGCTTTTGAACATGAAGAGCATGC 1007

 250 aAlaAlaPheSerLeuValLeuAspArgLeuArgGluThrGluAsnPheL 267
     |:::::|||||      ||||||||||||  :::|||   :::   :
1008 CAACACCTTTTCCGCGTTCCTTGACCGCCTTTCCGATACCGTCTCTGCAC 1057

 267 ysLysAspAlaGlyPheLysAlaGlnIleSerSerTrpLeuThrGlnLeu 283
     :::::   :::|||||:::::::|||:::::::::|||||   :::|||
1058 GCAATACCTCCGGATTCCGTGAACAGGTCGCTGCATGGCTGGAAAAACTC 1107

 284 AlaGluAspAlaAlaLeuArgAlaLysThrPheAlaMetAlaThrGluAl 300
     :::::::::|||:::|||||   :::::|||||||:::|||:::::::||
1108 AGTGCCTCTGCGGAGCTTCGACAGCAGTCTTTCGCTGTTGCTGCTGATGC 1157

 300 aThrSerThrCysGluAspArgValThrHisAlaLeuHisGlnMetAsnA 317
     ||||   :::|||||||||||||||:::   :::   :::::::::   :
1158 CACTGAGAGCTGTGAGGACCGTGTCGCGCTCACATGGAACAATCTCCGGA 1207

 317 snValGlnLeuValHisAsnAlaGluLysGlyGluTyrAspAsnAsnLeu 333
     :::::   ||||||||||:::|||     |||   :::|||||||:::
1208 AAACCCTCCTGGTCCATCAGGCATCAGAAGGCCTTTTCGATAATGATACC 1257

 334 GlnGlyLeuValSerThrGlyArgGluMetPheArgLeuAlaThrLeuGl 350
       :::|||:::|||   |||||||||||||||||||||||:::::|||||
1258 GGCGCTCTGCTCTCCCTGGGCAGGGAAATGTTCCGCCTCGAAATTCTGGA 1307

 350 uGlnIleAlaArgGluLysAlaGlyThrLeuAlaLeuValAspAspValG 367
     |:::|||||||||:::|||:::   ||||||   :::||||||:::::::|
1308 GGATATTGCCCGGGATAAAGTCAGAACTCTCCATTTTGTGGATGAGATAG 1357

 367 luValTyrLeuAlaPheGlnAsnLysLeuLysGluSerLeuGluLeuThr 383
     |||||||||||||||||||||:::::::|||   |||   |||:::|||:::
1358 AAGTCTACCTGGCCTTCCAGACCATGCTCGCAGAGAAACTTCAGCTCTCC 1407

 384 SerValThrSerGluMetArgPhePheAspValSerGlyValThrValSe 400
     :::::::::   ||||||||||||:::::::|||||||||||||||:::::
1408 ACTGCCGTGAAGGAAATGCGTTTCTATGGCGTGTCGGGAGTGACAGCAAA 1457

 400 rAspLeuGlnAlaAlaAspValGlnValLysThrAlaGluAsnSerGlyP 417
     :|||||:::::|||:::::::   |||::::::   |||::::::   |
1458 TGACCTCCGCACTGCCGAAGCCATGGTCAGAAGCCGTGAAGAGAATGAAT 1507

 417 heSerLysTrpIleLeuGlnTrpGlyProLeuHisSerValLeuGluArg 433
     ||:::   |||:::      |||||||||   |||:::||||||   |||
1508 TTACGGACTGGTTCTCCCTCTGGGGACCATGGCATGCTGTACTGAAGCGT 1557

 434 LysValProGluArgPheAsnAlaLeuArgGluLysGlnIleSerAspTy 450
     :::   ::::::|||   :::         |||::::::
1558 ACGGAAGCTGACCGCTGGGCGCAGGCAGAAGAGCAGAAATATGAGATGCT 1607

 450 rGluAspThrTyrArgLysLeuTyrAspGluValLeuLysSerSerGlyL 467
     |||:::   |||   :::      ::::::   ||||||:::|||||||
1608 GGAGAATGAGTACCCTCAGAGGGTGGCTGACCGGCTGAAAGCATCAGGTC 1657

 467 euValAspAspThrAspAlaGluArgThrIleGlyValSerAlaMetAsp 483
     ||   :::|||:::||||||||||||||      |||:::   :::|||
1658 TGAGCGGTGATGCGGATGCGGAGAGGGAAGCCGGTGCACAGGTGATGCGT 1707

 484 SerAlaLysLysGluPheLeuAspGlyLeuArgAlaLeuValAspGluVa 500
        :::   :::::::::          |||   |||:::|||||||||
1708 GAGACTGAACAGCAGATT.....TACCGTCAGCTGACTGACGAGGT 1748

 500 lLeuGlySerTyrLeuThrAlaArgTrpArgLeuAsn*** 513
     ||||:::                  ||||||:::|||
1749 ACTGGCC.................CTGCGATTGTCTGAA 1770
```

```
seq_name: gb_ba2:AE000074

seq_documentation_block:
LOCUS       AE000074    14263 bp    DNA             BCT       12-DEC-1997
DEFINITION  Rhizobium sp. NGR234 plasmid pNGR234a, section 11 of 46 of the
            complete plasmid sequence.
ACCESSION   AE000074 U00090
NID         g2182401
VERSION     AE000074.1  GI:2182401
KEYWORDS    .
SOURCE      Rhizobium sp. NGR234.
  ORGANISM  Rhizobium sp. NGR234
            Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
            Rhizobiaceae; Rhizobium.
REFERENCE   1  (bases 1 to 14263)
  AUTHORS   Freiberg,C., Fellay,R., Bairoch,A., Broughton,W.J., Rosenthal,A.
            and Perret,X.
  TITLE     Molecular basis of symbiosis between Rhizobium and legumes
  JOURNAL   Nature 387 (6631), 394-401 (1997)
  MEDLINE   97305956
REFERENCE   2  (bases 1 to 14263)
  AUTHORS   Freiberg,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-NOV-1996) Genome Analysis, Institute for Molecular
            Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
REFERENCE   3  (bases 1 to 14263)
  AUTHORS   Freiberg,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-DEC-1997) Genome Analysis, Institute for Molecular
            Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
  REMARK    Update by submitter
FEATURES             Location/Qualifiers
     source          1. .14263
                     /organism="Rhizobium sp. NGR234"
                     /plasmid="pNGR234a"
                     /strain="NGR234"
                     /db_xref="taxon:394"
     source          <1. .6634
                     /organism="Rhizobium sp. NGR234"
                     /db_xref="taxon:394"
                     /clone="pXBS23"
     source          <1. .>14263
                     /organism="Rhizobium sp. NGR234"
                     /db_xref="taxon:394"
                     /clone="pXBS4"
     misc_feature    131. .235
                     /note="mosaic sequence homolog (MSH) 14a: similar to
                     RIME1"
                     /note="Region: MSH14a"
     gene            548. .2188
                     /gene="y4fR"
     CDS             548. .2188
                     /gene="y4fR"
                     /note="hypothetical 60.5 Kd protein, homologous to
                     Shigella flexneri 60 Kd antigen (IpaH)"
                     /codon_start=1
                     /transl_table=11
                     /product="Y4fR"
                     /protein_id="AAB91674.1"
                     /db_xref="PID:g2182402"
```

```
                /db_xref="GI:2182402"
                /translation="MNVQRPGLAVGPLFENPESESSEPGSPAAAARWVEASTEAEASA
                ASSSQGQIVAAPTAEERPWEGRPQEAVSRTRAWREAGDVDEPLDLSFLSLTPLSIPLV
                SGLRRLNVNNNQLGDLPDTLPGTLLELEASENRLTRLPDLPAGLQRLNVENNRLTNLP
                EPLPAALEWLGAGYNQLTRLPEMIPPELIWLGARNNQLTSVPESLLTQLGQWSSIDLE
                NNPLPHGVQTNLVTAMHAAGYAGPQIFLPMGPVELARRPLHEVVADWLEGDLETVAAW
                RGFANEQGARDYAHFLDRLRTTVNYGNDAFRQAVAIGLRQAVARPQLRAQYFEQASGA
                SDSCEDRITLTWNGMQTALLIADVEDGVYDGSLHQLLQHGRVMFRLEALDGIARETVN
                SLRRTDPDADIDEIEVYLAYQTQLRDILELRHVAPDMRFLNVSHVTEEDVARAASSVR
                ELEARGFGEYVATRWQPWERVMRRIAPASHAAMQEQLIEAMGEEFRSRLDEKLAEHGL
                TGDADAERVFGAEILNDIARRIKGETMEKVLRGRGLEL"
gene            complement(2541. .3104)
                /gene="y4gA"
CDS             complement(2541. .3104)
                /gene="y4gA"
                /note="hypothetical 20.9 Kd protein; low similarity to
                Y4rE"
                /codon_start=1
                /transl_table=11
                /product="Y4gA"
                /protein_id="AAB91675.1"
                /db_xref="PID:g2182403"
                /db_xref="GI:2182403"
                /translation="MCQWAATKGTRNHAYEREKLERAFAKILRHRDPSVPPRRSDPRP
                RKEVARQWRKPHIYSPADLRRMLDIARSYPSPRAALRQENMYTMLLLAYCAGLRRGEL
                ARLDLGDVNLGDGTITVRQTKFFKTRILPLPDSVVVSFGPISLHGVGSAHSRLLTPPC
                SGTNKAVPAPRQEGSPGCLQTSYVVPG"
misc_feature    2541. .3104
                /note="insertion sequence homolog (ISH) 12f: region
                including a putative IS element-derived ORF (y4gA)"
                /note="Region: ISH12f"
gene            complement(3220. .3639)
                /gene="y4gB"
CDS             complement(3220. .3639)
                /gene="y4gB"
                /note="hypothetical 16.1 Kd protein"
                /codon_start=1
                /transl_table=11
                /product="Y4gB"
                /protein_id="AAB91676.1"
                /db_xref="PID:g2182404"
                /db_xref="GI:2182404"
                /translation="MTARWPDPDRAYIGRYVASLDLRSIKSRTCYRQVLHGFQDVVER
                HEALDQQVLLAWLRQSSDRWAATTLLHRTRIIDRFLDHFLEAAAIDHNPVEDLREACH
                IKQCMPVWRALISCDPEQALAQLRQPEPFGSAGRDHG"
misc_feature    complement(3635. .4215)
                /note="insertion sequence homolog (ISH) 12e: region
                including an integrase/recombinase (y4gC)"
                /note="Region: ISH12e"
gene            complement(3636. .4214)
                /gene="y4gC"
CDS             complement(3636. .4214)
                /gene="y4gC"
                /note="putative integrase/recombinase"
                /codon_start=1
                /transl_table=11
                /product="Y4gC"
                /protein_id="AAB91677.1"
                /db_xref="PID:g2182405"
                /db_xref="GI:2182405"
                /translation="MPSILERDQIAAVLVSASRDKTPAGLRDHAILQLLATYGLRSGE
                IRNMRIEDIDWRTETIRVRHSKTQAYTLLPLTEPVGEAILVYLRSGRPATDARELFVR
                TRAPYRKLDKLYSMVRRRLRDAGVKPRGKRGPHIFRHARATELLRAAVPQKIIGELLG
                HRSIASTAPYLKLATEDLRAIALDLPGTEVSA"
source          3683. .>14263
                /organism="Rhizobium sp. NGR234"
                /db_xref="taxon:394"
                /clone="pXB285"
gene            complement(4314. .4592)
                /gene="y4gD"
CDS             complement(4314. .4592)
                /gene="y4gD"
                /note="hypothetical 10.5 Kd protein"
                /codon_start=1
                /transl_table=11
                /product="Y4gD"
                /protein_id="AAB91678.1"
                /db_xref="PID:g2182406"
                /db_xref="GI:2182406"
                /translation="MQDRWPPAPKVTCAADALRFAICNEYETWLREERGLARASIDAF
                LWEARHFLGWQLERCGAEGLMEVSVGDIDRYVDMLASKLTRSSLKCAA"
misc_feature    5350. .5398
                /note="nod box-like sequence"
                /note="Region: nod box 5"
                /evidence=not_experimental
misc_feature    6735. .7535
                /note="insertion sequence homolog (ISH) 15"
                /note="Region: ISH15"
gene            6756. .7502
                /gene="y4gE"
CDS             6756. .7502
                /gene="y4gE"
                /note="hypothetical 27.7 Kd protein (possibly
                fragmentous)"
                /codon_start=1
                /transl_table=11
                /product="Y4gE"
                /protein_id="AAB91679.1"
                /db_xref="PID:g2182407"
                /db_xref="GI:2182407"
                /translation="MRAGQGGSLRLSVCSVIVQAPSRRALDAKARHAVRDDTILRQLK
                RDAAPRSLIATPQARQVADRFHLYADLRVAIEEQMSLSGRARGRALLPDKIIGNAQVD
                LIQDDPHVDATHRRRVRHGHRQSRQVVFETVHALRKKGLSCSAIARRTGYGRRSIAKW
                LTFETPPDRRKSVLKPTSPIKDGNRCGRHLYTISNSAVTSAVSRISSDFSQPGAAPRG
                RARTVRRRLLSYPINRLAMLSRCGIRRLAT"
gene            8469. .9521
                /gene="y4gF"
CDS             8469. .9521
                /gene="y4gF"
                /note="probable dTDP-D-glucose-4,6-dehydratase"
                /codon_start=1
                /transl_table=11
                /product="Y4gF"
                /protein_id="AAB91680.1"
                /db_xref="PID:g2182408"
                /db_xref="GI:2182408"
                /translation="MRILVTGGAGFIGSALVRYLVSINAEVLNVDKLTYAGNLASLKP
                VEGLRNYRFLRADICDRVAINEAFETFQPDYVIHLAAESHVDRSITGADDFVQTNVNG
                TFTMLETARQYWSNLSQNRKAFFKMLHVSTDEVYGSLGDRGQFEEVSPYDPSSPYSAS
                KAASDHFATAWQRTYGLPVVISNCSNNYGPFHFPEKLIPLMILNALDRKPLPVYGTGS
                NIRDWLYVDDHARALWLIVREGRPGEKYNVGGRNELRNIDVVNRICLLLDELSPNASH
                YGDLITFVKDRPGHDARYAIDATKLETELGWKAQENFDTGIRKTVEWYLENGWWWQPL
                RDKVYSGERLGLLEKA"
gene            9525. .10415
                /gene="y4gG"
CDS             9525. .10415
                /gene="y4gG"
                /note="probable dTDP-4-dehydrorhamnose reductase"
                /codon_start=1
                /transl_table=11
                /product="Y4gG"
                /protein_id="AAB91681.1"
                /db_xref="PID:g2182409"
                /db_xref="GI:2182409"
                /translation="MRLAVTGKNGQIALALKAQARPDVEILTLGRPNFDLACRSTVAS
                SIRDAAPDIIVSLAAYTAVDKAESEPYEAFAVNRDGVQALAEAAAGLGVPVIHLSTDY
                VFDGAKPVPYCEEDRTGPISVYGRSKLEGEFAVASANPNHTILRTSWVYSRYGQNFVK
                KMLRLADTNDELNVVADQLGCPTSADDISVAVMTIARRMLSSSSADLRGIFHLSGSGE
                ASWAAFAKYVFSVYDEITGRQIKVHDISAAEYPTPARRPANSRLHCDKLERTFGIRLP
                NWEESTRRLVWALLLEGKDA"
gene            10412. .11272
                /gene="y4gH"
CDS             10412. .11272

alignment_scores:
        Quality:  619.00          Length:    444
```

```
            Ratio:  2.063                Gaps:     12
Percent Similarity:  67.568   Percent Identity: 37.838

alignment_block:
US-09-068-804-14 x AE000074   ..

Align seg 1/1  to: AE000074  from: 1  to: 14263

     59 LeuGlnLeuThrSerLeuPro...SerLeuProGlnGlyLeuGlnLysLe 74
821 CTATCCCTGACCCCCTTGTCCATCCCCCTCGTATCCGGGCTCCGGCGCCT 870

     74 uTrpAlaTyrAsnAsnTrpLeuAlaSerLeuPro...ThrLeuProProG 90
871 GAACGTCAACAACAATCAGCTAGGGGATCTGCCAGACACCCTCCCGGGAA 920

     90 lyLeuGlyAspLeuAlaValSerAsnAsnGlnLeuThrSerLeuProGlu 106
921 CGCTTCTGGAGCTGGAGGCCAGCGAGAACCGGCTGACCCGCCTGCCCGAC 970

    107 MetProProAlaLeuArgGluLeuArgValSerGlyAsnAsnLeuThrSe 123
971 CTTCCTGCCGGGCTCCAGAGGCTGAACGTCGAAAACAACCGGCTGACCAA 1020

    123 rLeuArg...AlaLeuProSerGlyLeuGlnLysLeuTrpAlaTyrAsnA 139
1021 TCTCCCCGAGCCCCTCCCGGCCGCGCTCGAGTGGCTGGGTGCTGGCTACA 1070

    139 snArgLeuThrSerLeuProGluMet...SerProGlyLeuGlnGluLeu 154
1071 ACCAGCTAACCCGCCTGCCCGAGATGATTCCCCCCGAGCTTATATGGCTG 1120

    155 AspValSerHisAsnGlnLeuThrArgLeuProGlnSer...LeuThrGl 170
1121 GGCGCCAGAAACAATCAGCTGACCAGCGTGCCCGAAAGCCTGCTGACGCA 1170

    170 yLeuSerSerAlaAlaArgValTyrLeuAspGlyAsnProLeuSerValA 187
1171 GCTAGGCCAATGGTCTAGCATTGATTTGGAAAATAATCCGCTGCCGCATG 1220

    187 rgThrArgAspArgLeuCysGlyHisHisTrpProPheArgHisGlnAsp 203
1221 GGGTACAAACAAACTTGGTGACA.................GCCATGCAT 1252

    204 ThrLeuArgTyrGlyGlyAla......PheArgProArgGluAla..... 216
1253 GCCGCGGGTTACGCAGGACCGCAGATCTTTTTGCCTATGGGACCGGTGGA 1302

    217 ..........ArgAlaLeuHisLeuAlaValAlaAspTrpLeuThrSerA 230
1303 ACTTGCACGACGGCCCCTTCACGAGGTTGTCGCGGATTGGCTC....... 1345

    230 laArgGluGlyGluAlaAlaGlnAlaAspArgTrpGlnAlaPheGlyLeu 246
1346 .....GAGGGGGATTTGGAAACGGTGGCCGCCTGGCGGGGCTTCGCCAAC 1390

    247 GluAspAsnAlaAlaAlaPheSerLeuValLeuAspArgLeuArgGluTh 263
1391 GAACAAGGTGCCCGGGATTATGCACATTTCCTCGACCGGCTCCGTACCAC 1440

    263 rGluAsnPheLysLysAspAlaGlyPheLysAlaGlnIleSerSerTrpL 280
1441 CGTGAACTATGGCAATGATGCG...TTCCGGCAGGCGGTGGCCATAGGTC 1487

    280 euThrGlnLeuAlaGluAspAlaAlaLeuArgAlaLysThrPheAlaMet 296
1488 TGCGGCAGGCAGTGGCCAGGCCGCAATTGCGCGCACAGTATTTTGAGCAG 1537

    297 AlaThrGluAlaThrSerThrCysGluAspArgValThrHisAlaLeuHi 313
1538 GCTTCCGGAGCTAGCGATAGCTGTGAGGATCGCATTACTTTGACCTGGAA 1587

    313 sGlnMetAsnAsnValGlnLeuValHisAsnAlaGluLysGlyGluTyrA 330
1588 TGGAATGCAGACCGCACTTCTGATCGCTGATGTAGAGGACGGGGTCTATG 1637

    330 spAsnAsnLeuGlnGlyLeuValSerThrGlyArgGluMetPheArgLeu 346
1638 ACGGGTCGCTCCACCAGCTTCTCCAGCACGGCCGTGTCATGTTCCGTTTG 1687

    347 AlaThrLeuGluGlnIleAlaArgGluLysAlaGlyThrLeu..... 360
1688 GAGGCACTGGACGGGATCGCGCGCGAGACGGTCAACTCGCTTCGACGTAC 1737

    361 .....AlaLeuValAspAspValGluValTyrLeuAlaPheGlnA 374
1738 CGATCCGGACGCCGACATCGACGAGATCGAGGTCTATCTTGCCTACCAGA 1787

    374 snLysLeuLysGluSerLeuGluLeuThrSerValThrSerGluMetArg 390
1788 CACAGCTGCGCGACACCCTGGAGCTGCGGCACGTCGCCCCAGACATGCGC 1837

    391 PhePheAspValSerGlyValThrValSerAspLeuGlnAlaAlaAspVa 407
1838 TTCTTGAACGTCTCTCACGTTACCGAGGAGGACGTTGCCAGAGCTGCCAG 1887

    407 lGlnValLysThrAlaGluAsnSerGlyPheSerLysTrpIle...LeuG 423
1888 TTCCGTGCGGGAACTTGAGGCGAGGGGATTCGGCGAGTATGTGGCAACGC 1937

    423 lnTrpGlyProLeuHisSerValLeuGluArgLysValProGluArgPhe 439
1938 GCTGGCAACCTTGGGAAAGAGTCATGAGGCGCATCGCCCCCGCAAGCCAT 1987

    440 AsnAlaLeuArgGluLysGlnIleSerAspTyrGluAspThrTyrArgLy 456
1988 GCAGCGATGCAGGAGCAGCTCATCGAAGCGATGGGAGAGGAGTTCAGAAG 2037

    456 sLeuTyrAspGluValLeuLysSerSerGlyLeuValAspAspThrAspA 473
2038 CCGCTTGGACGAAAAGCTCGCCGAGCACGGCCTGACGGGTGATGCCGATG 2087

    473 laGluArgThrIleGlyValSerAlaMetAsp 483
2088 CCGAACGGGTATTCGGAGCCGAGATCCTCAAC 2119

seq_name: gb_bal:SHFIPAHX

seq_documentation_block:
LOCUS       SHFIPAHX     1312 bp    DNA             BCT       23-OCT-1991
DEFINITION  Shigella flexneri ipaH gene.
ACCESSION   M76443
NID         g152749
VERSION     M76443.1  GI:152749
KEYWORDS    LRG-like protein; ipaH gene.
SOURCE      Shigella flexneri (strain 5) DNA.
  ORGANISM  Shigella flexneri
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Shigella.
REFERENCE   1  (bases 1 to 1312)
  AUTHORS   Venkatesan,M.M., Buysse,J.M. and Hartman,A.B.
  TITLE     Sequence variation in two ipaH genes of Shigella flexneri 5 and
            homology to the LRG-like family of proteins
  JOURNAL   Mol. Microbiol. 5, 2435-2446 (1991)
  MEDLINE   92167809
FEATURES             Location/Qualifiers
     source          1. .1312
                     /organism="Shigella flexneri"
                     /strain="5"
                     /db_xref="taxon:623"
BASE COUNT      326 a    328 c    338 g    320 t
ORIGIN


alignment_scores:
            Quality:  602.50              Length:    388
```

```
            Ratio:  2.257                Gaps:      7
 Percent Similarity:  68.814   Percent Identity: 36.856

alignment_block:
 US-09-068-804-14 x SHFIPAHX   ..

 Align seg 1/1  to: SHFIPAHX  from: 1  to: 1312

     144 LeuProGluMetSerProGlyLeuGlnGluLeuAspValSerHisAsnGl 160
         |||||||||:::::   :::   :::          :::   :::|||::
      33 CTACCGGAATTACCTTTTAGCTCGAAACGGGCTGTGCTAATGAATAATAA 82

     160 nLeuThrArgLeuProGlnSerLeuThrGlyLeuSerSerAlaAlaArgV 177
         :||||||   ||||||:::|||:::      |||:::   :::|||   |
      83 TCTGACAACACTTCCGGAAAGTGTCCTGAGATTAGCTCAGAATGCCTTCG 132

     177 alTyrLeuAspGlyAsnProLeuSerValArgThrArgAspArgLeuCys 193
         ||   :::::|||||||||||||
     133 TAAATGTTGCAGGTAATCCATTGTC..................................... 157

     194 GlyHisHisTrpProPheArg.............HisGlnAspThrLe 205
                 ||||||:::::              :::::::::
     158 .........TGGCCATACCATGCGTACACTACAACAAATAACCACCGGAC 198

     205 uArgTyrGlyGlyAlaPheArgProArg..................... 214
             |||:::|||         ||||||
     199 CAGTTATTCTGGT.........CCTCGAATATTTTTCTCTATGGGAAATT 239

     215 .................GluAlaArgAlaLeuHisLeuAlaValAlaAsp 225
                         ::::::::::|||      ||||||:::::
     240 CTGCCACAATTTCCGCTCCAGAACACTCCCTGGCTGATGCCGTGACAGCA 289

     226 TrpLeuThrSerAlaArgGluGlyGluAlaAlaGlnAlaAspArgTrpGl 242
         |||:::::::   ::::::::::::::::::::::|||         |||::
     290 TGGTTCCCGGAAAACAAACAATCTGATGTATCACAGATA......TGGCA 333

     242 nAlaPheGlyLeuGluAspAsnAlaAlaAlaPheSerLeuValLeuAspA 259
         :|||||||      |||:::::::|||::::::|||||||      |||||||
     334 TGCTTTTGAACATGAAGAGCATGCCAACACCTTTTCCGCGTTCCTTGACC 383

     259 rgLeuArgGluThrGluAsnPheLysLysAspAlaGlyPheLysAlaGln 275
         |||||   :::|||   :::   ::::::   :::||||||::::::|||
     384 GCCTTTCCGATACCGTCTCTGCACGCAATACCTCCGGATTCCGTGAACAG 433

     276 IleSerSerTrpLeuThrGlnLeuAlaGluAspAlaAlaLeuArgAlaLy 292
         :::::::::||||||   :::|||:::::::::|||:::||||||   ::
     434 GTCGCTGCATGGCTGGAAAAACTCAGTGCCTCTGCGGAGCTTCGACAGCA 483

     292 sThrPheAlaMetAlaThrGluAlaThrSerThrCysGluAspArgValT 309
         ::::||||||:::|||:::::::||||||   :::|||||||||||||||:
     484 GTCTTTCGCTGTTGCTGCTGATGCCACTGAGAGCTGTGAGGACCGTGTCG 533

     309 hrHisAlaLeuHisGlnMetAsnAsnValGlnLeuValHisAsnAlaGlu 325
         ::   :::   :::::::::   ::::::   |||||||||:::|||
     534 CGCTCACATGGAACAATCTCCGGAAAACCCTCCTGGTCCATCAGGCATCA 583

     326 LysGlyGluTyrAspAsnAsnLeuGlnGlyLeuValSerThrGlyArgGl 342
            |||   :::|||||||:::      :::|||:::|||   ||||||||
     584 GAAGGCCTTTTCGATAATGATACCGGCGCTCTGCTCTCCCTGGGCAGGGA 633

     342 uMetPheArgLeuAlaThrLeuGluGlnIleAlaArgGluLysAlaGlyT 359
         |||||||||||||:::::|||||||:::|||||||||:::|||:::   |
     634 AATGTTCCGCCTCGAAATTCTGGAGGACATTGCCCGGGATAAAGTCAGAA 683

     359 hrLeuAlaLeuValAspAspValGluValTyrLeuAlaPheGlnAsnLys 375
         |||||   :::|||||||:::::::|||||||||||||||||||:::::::
     684 CTCTCCATTTTGTGGATGAGATAGAAGTCTACCTGGCCTTCCAGACCATG 733

     376 LeuLysGluSerLeuGluLeuThrSerValThrSerGluMetArgPhePh 392
         |||   |||   |||:::|||:::::::::::::   ||||||||||||::
     734 CTCGCAGAGAAACTTCAGCTCTCCACTGCCGTGAAGGAAATGCGTTTCTA 783

     392 eAspValSerGlyValThrValSerAspLeuGlnAlaAlaAspValGlnV 409
         :::||||||||||||||||||:::::||||||:::::::|||:::::   |
     784 TGGCGTGTCGGGAGTGACAGCAAATGACCTCCGCACTGCCGAAGCTATGG 833

     409 alLysThrAlaGluAsnSerGlyPheSerLysTrpIleLeuGlnTrpGly 425
         ||:::::   |||:::::   |||:::   |||:::      ||||||
     834 TCAGAAGCCGTGAAGAGAATGAATTTACGGACTGGTTCTCCCTCTGGGGA 883

     426 ProLeuHisSerValLeuGluArgLysValProGluArgPheAsnAlaLe 442
         |||   |||:::||||||   |||:::   :::::|||   :::
     884 CCATGGCATGCTGTACTGAAGCGTACGGAAGCTGACCGCTGGGCGCAGGC 933

     442 uArgGluLysGlnIleSerAspTyrGluAspThrTyrArgLysLeuTyrA 459
             |||::::::            |||:::   |||   :::      :
     934 AGAAGAGCAGAAGTATGAGATGCTGGAGAATGAGTACTCTCAGAGGGTGG 983

     459 spGluValLeuLysSerSerGlyLeuValAspAspThrAspAlaGluArg 475
         :::::   ||||||:::|||||||||   :::|||:::||||||:::|||
     984 CTGACCGGCTGAAAGCATCAGGTCTGAGCGGTGATGCGGATGCGCAGAGG 1033

     476 ThrIleGlyValSerAlaMetAspSerAlaLysLysGluPheLeuAspGl 492
             |||:::   :::|||      :::   :::::::::
    1034 GAAGCCGGTGCACAGGTGATGCGTGAGACTGAACAGCAGATT........ 1075

     492 yLeuArgAlaLeuValAspGluValLeuGlySerTyrLeuThrAlaArgT 509
         |||   |||:::|||||||||||||:::
    1076 .TACCGTCAGCTGACTGACGAGGTACTGGCC.................C 1106

     509 rpArgLeuAsn*** 513
         ||||||:::|||
    1107 TGCGATTGTCTGAA 1120

seq_name: gb_ba1:SHFIPAHY

seq_documentation_block:
LOCUS       SHFIPAHY     1375 bp    DNA              BCT       23-OCT-1991
DEFINITION  Shigella flexneri ipaH gene.
ACCESSION   M76444
NID         g152750
VERSION     M76444.1  GI:152750
KEYWORDS    LRG-like protein; ipaH gene.
SOURCE      Shigella flexneri (strain 5) DNA.
  ORGANISM  Shigella flexneri
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Shigella.
REFERENCE   1  (bases 1 to 1375)
  AUTHORS   Venkatesan,M.M., Buysse,J.M. and Hartman,A.B.
  TITLE     Sequence variation in two ipaH genes of Shigella flexneri 5 and
            homology to the LRG-like family of proteins
  JOURNAL   Mol. Microbiol. 5, 2435-2446 (1991)
  MEDLINE   92167809
FEATURES             Location/Qualifiers
     source          1. .1375
                     /organism="Shigella flexneri"
                     /strain="5"
                     /db_xref="taxon:623"
BASE COUNT      343 a    347 c    370 g    315 t
ORIGIN


alignment_scores:
            Quality:  602.50              Length:    388
              Ratio:  2.248                 Gaps:      7
 Percent Similarity:  69.072   Percent Identity: 37.113

alignment_block:
 US-09-068-804-14 x SHFIPAHY   ..

 Align seg 1/1  to: SHFIPAHY  from: 1  to: 1375

     144 LeuProGluMetSerProGlyLeuGlnGluLeuAspValSerHisAsnGl 160
         |||||||||:::::   :::   :::          :::   :::|||::
      33 CTACCGGAATTACCTTTTAGCTCGAAACGGGCTGTGCTAATGAATAATAA 82
```

```
160 nLeuThrArgLeuProGlnSerLeuThrGlyLeuSerSerAlaAlaArgV 177
 83 TCTGACAACACTTCCGGAAAGTGTCCTGAGATTAGCTCAGAATGCCTTCG 132

177 alTyrLeuAspGlyAsnProLeuSerValArgThrArgAspArgLeuCys 193
133 TAAATGTTGCAGGTAATCCATTGTC..........................  157

194 GlyHisHisTrpProPheArg..............HisGlnAspThrLe 205
158 .........TGGCCATACCATGCGTACACTACAACAAATAACCACCGGAC 198

205 uArgTyrGlyGlyAlaPheArgProArg..................... 214
199 CAGTTATTCTGGT.........CCTCGAATATTTTTCTCTATGGGAAATT 239

215 .................GluAlaArgAlaLeuHisLeuAlaValAlaAsp 225
240 CTGCCACAATTTCCGCTCCAGAACACTCCCTGGCTGATGCCGTGACAGCA 289

226 TrpLeuThrSerAlaArgGluGlyGluAlaAlaGlnAlaAspArgTrpGl 242
290 TGGTTCCCGGAAAACAAACAATCTGATGTATCACAGATA......TGGCA 333

242 nAlaPheGlyLeuGluAspAsnAlaAlaAlaPheSerLeuValLeuAspA 259
334 TGCTTTTGAACATGAAGAGCATGCCAACACCTTTTCCGCGTTCCTTGACC 383

259 rgLeuArgGluThrGluAsnPheLysLysAspAlaGlyPheLysAlaGln 275
384 GCCTTTCCGATACCGTCTCTGCACGCAATACCTCCGGATTCCGTGAACAG 433

276 IleSerSerTrpLeuThrGlnLeuAlaGluAspAlaAlaLeuArgAlaLy 292
434 GTCGCTGCATGGCTGGAAAAACTCAGTGCCTCTGCGGAGCTTCGACAGCA 483

292 sThrPheAlaMetAlaThrGluAlaThrSerThrCysGluAspArgValT 309
484 GTCTTTCGCTGTTGCTGCTGATGCCACTGAGAGCTGTGAGGACCGTGTCG 533

309 hrHisAlaLeuHisGlnMetAsnAsnValGlnLeuValHisAsnAlaGlu 325
534 CGCTCACATGGAACAATCTCCGGAAAACCCTCCTGGTCCATCAGGCATCA 583

326 LysGlyGluTyrAspAsnAsnLeuGlnGlyLeuValSerThrGlyArgGl 342
584 GAAGGCCTTTTCGATAATGATACCGGCGCTCTGCTCTCCCTGGGCAGGGA 633

342 uMetPheArgLeuAlaThrLeuGluGlnIleAlaArgGluLysAlaGlyT 359
634 AATGTTCCGCCTCGAAATTCTGGAGGACATTGCCCGGGATAAAGTCAGAA 683

359 hrLeuAlaLeuValAspAspValGluValTyrLeuAlaPheGlnAsnLys 375
684 CTCTCCATTTTGTGGATGAGATAGAAGTCTACCTGGCCTTCCAGACCATG 733

376 LeuLysGluSerLeuGluLeuThrSerValThrSerGluMetArgPhePh 392
734 CTCGCAGAGAAACTTCAGCTCTCCACTGCCGTGAAGGAAATGCGTTTCTA 783

392 eAspValSerGlyValThrValSerAspLeuGlnAlaAlaAspValGlnV 409
784 TGGCGTGTCGGGAGTGACAGCAAATGACCTCCGCACTGCCGAAGCTATGG 833

409 alLysThrAlaGluAsnSerGlyPheSerLysTrpIleLeuGlnTrpGly 425
834 TCAGAAGCCGTGAAGAGAATGAATTTACGGACTGGTTCTCCCTCTGGGGA 883

426 ProLeuHisSerValLeuGluArgLysValProGluArgPheAsnAlaLe 442
884 CCATGGCATGCTGTACTGAAGCGTACGGAAGCTGACCGCTGGGCGCAGGC 933

442 uArgGluLysGlnIleSerAspTyrGluAspThrTyrArgLysLeuTyrA 459

934 AGAAGAGCAGAAGTATGAGATGCTGGAGAATGAGTACTCTCAGAGGGTGG 983

459 spGluValLeuLysSerSerGlyLeuValAspAspThrAspAlaGluArg 475
984 CTGACCGGCTGAAAGCATCAGGTCTGAGCGGTGATGCGGATGCGCAGAGG 1033

476 ThrIleGlyValSerAlaMetAspSerAlaLysLysGluPheLeuAspGl 492
1034 GAAGCCGGTGCACAGGTGATGCGTGAGACTGAACAGCAGATT........ 1075

492 yLeuArgAlaLeuValAspGluValLeuGly.SerTyrLeuThrAlaArg 508
1076 .TACCGTCAGCTGACTGACGAGGTACTGGCCTGAACCGCCCCGGGAATCC 1124

509 TrpArgLeuAsn 512
1125 TGGAGACTAAAC 1136
```

```
seq_name: gb_ba1:YEPYOPM

seq_documentation_block:
LOCUS       YEPYOPM      1397 bp    DNA             BCT       21-JUN-1994
DEFINITION  Y.pestis outer membrane protein (yopM) gene, complete cds.
ACCESSION   M25810
NID         g155539
VERSION     M25810.1  GI:155539
KEYWORDS    outer membrane protein.
SOURCE      Y.pestis DNA.
  ORGANISM  Yersinia pestis
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Yersinia.
REFERENCE   1  (bases 1 to 1397)
  AUTHORS   Leung,K.Y. and Straley,S.C.
  TITLE     The yopM gene of Yersinia pestis encodes a released protein having
            homology with the human platelet surface protein GPIb-alpha
  JOURNAL   J. Bacteriol. 171, 4623-4632 (1989)
  MEDLINE   89359090
COMMENT     Draft entry and computer-readable sequence for [1] kindly submitted
            by S.C.Straley, 06-JUL-1989.
FEATURES             Location/Qualifiers
     source          1. .1397
                     /organism="Yersinia pestis"
                     /db_xref="taxon:632"
     CDS             164. .1267
                     /note="outer membrane protein (yopM)"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA27670.1"
                     /db_xref="PID:g155540"
                     /db_xref="GI:155540"
                     /translation="MFINPRNVSNTFLQEPLRHSSNLTEMPVEAENVKSKTEYYNAWS
                     EWERNAPPGNGEQREMAVSRLRDCLDRQAHELELNNLGLSSLPELPPHLESLVASCNS
                     LTELPELPQSLKSLLVDNNNLKALSDLPPLLEYLGVSNNQLEELPELQNLPFLTAIYA
                     DNNSLKKLPDLPLSLESIVAGNNILEELPELQNLPFLTTIYADNNLLKTLPDLPPSLE
                     ALNVRDNYLTDLPELPQSLTFLDVSENIFSGLSELPPNLYYLNASSNEIRSLCDLPPS
                     LEELNVSNNKLIELPALPPRLERLIASFNHLAEVPELPQNLKQLHVEYNPLREFPDIP
                     ESVEDLRMNSERVVDPYEFAHETTDKLEDDVFE"
BASE COUNT      450 a    274 c    257 g    416 t
ORIGIN


alignment_scores:
            Quality:  272.00              Length:    228
              Ratio:  2.109                Gaps:      4
 Percent Similarity:  56.579   Percent Identity: 34.211

alignment_block:
 US-09-068-804-14 x YEPYOPM   ..

 Align seg 1/1  to: YEPYOPM  from: 1  to: 1397

       1 MetArgAspCysLeuAsnAsnGlyAsnProValLeuAsnValGlyAlaSe 17
```

```
356 TTACGAGATTGCCTGGACCGACAAGCCCATGAGCTAGAACTAAATAATCT 405

 17 rGlyLeuThrThrLeuProAspArgLeuProProHisIleThrThrLeuV 34
406 GGGGCTGAGTTCTTTGCCGGAA...TTACCTCCGCATTTAGAGAGTTTAG 452

 34 alIleProAspAsnAsnLeuThrSerLeuProGluLeuProGluGlyLeu 50
453 TGGCGTCATGTAATTCTCTTACAGAATTACCGGAATTACCGCAGAGCCTG 502

 51 ArgGluLeuGluValSerGlyAsn..... 58
503 AAATCACTTCTAGTTGATAATAACAATCTGAAGGCATTATCCGATTTACC 552

 58 ..... 58
553 ACCTTTACTGGAATATTTAGGTGTCTCTAATAATCAGCTGGAAGAATTGC 602

 59 .....LeuGln..... 60
603 CAGAGTTGCAAAACTTGCCCTTCTTGACTGCGATTTATGCTGATAACAAT 652

 61 ...LeuThrSerLeuProSerLeuProGlnGlyLeuGln..... 72
653 TCACTGAAAAAACTACCTGATTTACCTCTTTCACTGGAATCTATTGTTGC 702

 72 ..... 72
703 TGGTAATAATATTCTGGAAGAATTGCCAGAGTTGCAAAACTTGCCCTTCT 752

 73 .....LysLeuTrpAlaTyrAsnAsnTrpLeuAlaSerLeuProThrLeu 87
753 TGACTACGATTTATGCTGATAACAATTTACTGAAAACATTACCCGATTTA 802

 88 ProProGlyLeuGlyAspLeuAlaValSerAsnAsnGlnLeuThrSerLe 104
803 CCCCCTTCCCTGGAAGCACTTAATGTCAGAGATAATTATTTAACTGATCT 852

104 uProGluMetProProAlaLeuArgGluLeuArgValSerGlyAsnAsnL 121
853 GCCAGAATTACCGCAGAGTTTAACCTTCTTAGATGTTTCTGAAAATATTT 902

121 euThrSerLeuArgAlaLeuProSerGlyLeuGlnLysLeuTrpAlaTyr 137
903 TTTCTGGATTATCGGAATTGCCACCAAACTTGTATTATCTCAATGCATCC 952

138 AsnAsnArgLeuThrSerLeuProGluMetSerProGlyLeuGlnGluLe 154
953 AGCAATGAAATAAGATCCTTATGCGATTTACCCCCTTCACTGGAAGAACT 1002

154 uAspValSerHisAsnGlnLeuThrArgLeuPro 165
1003 TAATGTCAGTAATAATAAGTTGATCGAACTGCCA 1036
```

seq_name: gb_ba2:AF102990

seq_documentation_block:

LOCUS AF102990 69673 bp DNA circular BCT 18-MAY-1999
DEFINITION Yersinia enterocolitica plasmid pYVe227, complete sequence.
ACCESSION AF102990 AF054978 AF054979 AF054980 AF054981 AF080156 Z69926 L06216 AF080155 AF022645 AF050104 AF054977 AF033863 U02499 U08019 U21297 M22781 U08222 Z18539 U94827
NID g4324323
VERSION AF102990.1 GI:4324323
KEYWORDS .
SOURCE Yersinia enterocolitica.
ORGANISM Yersinia enterocolitica
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Yersinia.
REFERENCE 1 (bases 32056 to 32871)
AUTHORS Cornelis,G., Sluiters,C., de Rouvroit,C.L. and Michiels,T.
TITLE Homology between virF, the transcriptional activator of the Yersinia virulence regulon, and AraC, the Escherichia coli arabinose operon regulator
JOURNAL J. Bacteriol. 171 (1), 254-262 (1989)
MEDLINE 89123026
REFERENCE 2 (bases 15074 to 15994)
AUTHORS Hakansson,S., Bergman,T., Vanooteghem,J.C., Cornelis,G. and Wolf-Watz,H.
TITLE YopB and YopD constitute a novel class of Yersinia Yop proteins
JOURNAL Infect. Immun. 61 (1), 71-80 (1993)
MEDLINE 93114907
REFERENCE 3 (bases 45325 to 45717)
AUTHORS Wattiau,P. and Cornelis,G.R.
TITLE SycE, a chaperone-like protein of Yersinia enterocolitica involved in Ohe secretion of YopE
JOURNAL Mol. Microbiol. 8 (1), 123-131 (1993)
MEDLINE 93268087
REFERENCE 4 (bases 23950 to 25269)
AUTHORS Woestyn,S., Allaoui,A., Wattiau,P. and Cornelis,G.R.
TITLE YscN, the putative energizer of the Yersinia Yop secretion machinery
JOURNAL J. Bacteriol. 176 (6), 1561-1569 (1994)
MEDLINE 94179088
REFERENCE 5 (bases 29897 to 30961)
AUTHORS Allaoui,A., Woestyn,S., Sluiters,C. and Cornelis,G.R.
TITLE YscU, a Yersinia enterocolitica inner membrane protein involved in Yop secretion
JOURNAL J. Bacteriol. 176 (15), 4534-4542 (1994)
MEDLINE 94321323
REFERENCE 6 (bases 43292 to 43717)
AUTHORS Wattiau,P., Bernier,B., Deslee,P., Michiels,T. and Cornelis,G.R.
TITLE Individual chaperones required for Yop secretion by Yersinia
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 91 (22), 10493-10497 (1994)
MEDLINE 95024141
REFERENCE 7 (bases 31537 to 31932)
AUTHORS Allaoui,A., Scheen,R., Lambert de Rouvroit,C. and Cornelis,G.R.
TITLE VirG, a Yersinia enterocolitica lipoprotein involved in Ca2+ dependency, is related to exsB of Pseudomonas aeruginosa
JOURNAL J. Bacteriol. 177 (15), 4230-4237 (1995)
MEDLINE 95362644
REFERENCE 8 (bases 53135 to 53497)
AUTHORS Stainier,I., Iriarte,M. and Cornelis,G.R.
TITLE YscM1 and YscM2, two Yersinia enterocolitica proteins causing downregulation of yop transcription
JOURNAL Mol. Microbiol. 26 (4), 833-843 (1997)
MEDLINE 98087284
REFERENCE 9 (bases 18681 to 18968)
AUTHORS Boyd,A.P., Sory,M.P., Iriarte,M. and Cornelis,G.R.
TITLE Heparin interferes with translocation of Yop proteins into HeLa cells and binds to LcrG, a regulatory component of the Yersinia Yop apparatus
JOURNAL Mol. Microbiol. 27 (2), 425-436 (1998)
MEDLINE 98143428
REFERENCE 10 (bases 22612 to 22890)
AUTHORS Iriarte,M., Sory,M.P., Boland,A., Boyd,A.P., Mills,S.D., Lambermont,I. and Cornelis,G.R.
TITLE TyeA, a protein involved in control of Yop release and in translocation of Yersinia Yop effectors
JOURNAL EMBO J. 17 (7), 1907-1918 (1998)
MEDLINE 98190073
REFERENCE 11 (bases 1 to 69673)
AUTHORS Iriarte,M., Lambermont,I., Kerbourch,C. and Cornelis,G.R.
TITLE Detailed genetic map of the pYVe227 plasmid of Yersinia enterocolitica serotype O:9
JOURNAL Unpublished
REFERENCE 12 (bases 1 to 69673)
AUTHORS Iriarte,M., Lambermont,I., Kerbourch,C. and Cornelis,G.R.
TITLE Direct Submission
JOURNAL Submitted (30-OCT-1998) Microbial Pathogenesis Unit, Christian de Duve Institute of Cellular Pathology and Faculte de Medecine. Universite Catholique de Louvain, Av. Hippocrate 74.49, Brussels 1200, Belgium
COMMENT On May 18, 1999 this sequence version replaced gi:3411136 gi:3411135 gi:3411139 gi:3411137 gi:3603360 gi:1524380 gi:155536

```
          gi:3603357 gi:2738986 gi:2944089 gi:3411138 gi:3108216 gi:437201
          gi:497215 gi:885971 gi:155526 gi:483322 gi:48593 gi:2735421.
FEATURES             Location/Qualifiers
     source          1. .69673
                     /organism="Yersinia enterocolitica"
                     /strain="W22703"
                     /serotype="O:9"
                     /db_xref="taxon:630"
                     /plasmid="pYVe227"
     rep_origin      53. .203
     CDS             complement(889. .1434)
                     /note="ORF181"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAD16801.1"
                     /db_xref="PID:g4324324"
                     /db_xref="GI:4324324"
                     /translation="MRVVIYQGQYGIFKKNGLDLSMPAIQTCLGLYALSDQHDYLLCA
                     HFDSAFGLQQNLQDIKRFIELNGIRMSSLRATVFGGDGKQSYLRCSAPSSHIGNEIVN
                     FMKSHGALAQYSNQYYSGVFPETFNFHYQKGCGITKGQNPRDFLGGSPQAMNLAKQRI
                     KLRPSEYSLTHAKMLDISRLY"
     CDS             1616. .2083
                     /note="ORF155"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAD16802.1"
                     /db_xref="PID:g4324325"
                     /db_xref="GI:4324325"
                     /translation="MINTTFTELLQKIASHFGLDKLSQDEYGLCELILNDRVVIMLRA
                     DEILNRLILLGPILGFSGPEARSTASQLFFCYSINALNKDGPCFAWSEELGLIAFKHL
                     SLGELNVENVSKEIANFYDWLSLVSLPAETQQELPPHTQSTQSTQSTQSVKWG"
     gene            2091. .4280
                     /gene="yopO"
     CDS             2091. .4280
                     /gene="yopO"
                     /codon_start=1
                     /transl_table=11
                     /product="protein kinase YopO"
                     /protein_id="AAD16803.1"
                     /db_xref="PID:g4324326"
                     /db_xref="GI:4324326"
                     /translation="MKIMGTMPPSISLAKAHERISQHWQNPVGELNIGGKRYRIIDNQ
                     VLRLNPHSGFSLFREGVGKIFSGKMFNFSIARNLTETLHAAQKTTSQELRSDIPNALS
                     NLFGAKPQTELPLGWKGKPLSGAPDLEGMRVAETDKFAEGESHISIIETKDNQRLVAK
                     IERSIAEGHLFAELEAYKHIYKTAGKHPNLANVHGMAVVPYGNRKEEALLMDEVDGWR
                     CSDTLRSLADSWKQGKINSEAYWGTIKFIAHRLLDVTNHLAKAGIVHNDIKPGNVVFD
                     RASGEPVVIDLGLHSRSGEQPKGFTESFKAPELGVGNLGASEKSDVFLVVSTLLHGIE
                     GFEKDPEIKPNQGLRFITSEPAHVMDENGYPIHRPGIAGVETAYTRFITDILGVSADS
                     RPDSNEARLHEFLSDGTIDEESAKQILKDTLTGEMSPLSTDVRRITPKKLRELSDLLR
                     THLSSAATKQLDMGVVLSDLDTMLVTLDKAEREGGVDKDQLKSFNSLILKTYSVIEDY
                     VKGREGDTKSSSAEVSPYHRSNFMLSIAEPSLQRIQKHLDQTHSFSDIGSLVRAHKHL
                     ETLLEVLVTLSPQGQPVSSETYSFLNRLAEAKVTLSQQLDTLQQQQESAKAQLSILIN
                     RSGSWADVARQSLQRFDSTRPVVKFGTEQYTAIHRQMMAAHAAITLQEVSEFTDDMRN
                     FTADSIPLLIRLGRSSLIDEHLVEQREKLRELTTIAERLNRLEREWM"
     CDS             4485. .4760
                     /note="ORF91A; unknown function"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAD16804.1"
                     /db_xref="PID:g4324327"
                     /db_xref="GI:4324327"
                     /translation="MLSIRKNECPCWTLNNGLPVICSGELAINEGKFVGINNYSALYR
                     IMNLGVLVRVLPVCSGFTHQRLNVLLQGTVYPEYKIHTAVNSLNKDK"
     gene            4761. .5627
                     /gene="yopP"
     CDS             4761. .5627
                     /gene="yopP"
                     /codon_start=1
                     /transl_table=11
                     /product="Yop effector YopP"
                     /protein_id="AAD16805.1"
                     /db_xref="PID:g4324328"
                     /db_xref="GI:4324328"
                     /translation="MIGPISQINSFGGLSEKETRSLISNEELKNIIIQLETDIADGSW
                     FHKNYSRLDIEVMPALVIQANNKYPEMNLNFVTSPQDLSIEIKNVIENGVGSSRFIIN
                     MGEGGIHFSVIDYKHINGKTSLILFEPVNFNSMGPAILAISTKTAIERYQLPDCHFSM
                     VEMDIQRSSSECGIFSLALAKKLYTERDSLLKIHEDNIKGILSDSENPLPHNKLDPYL
                     PVTFYKHTQGKKRLNEYLNTNPQGVGTVVNKKNETIFNRFDNNKSIIDGKELSVSVHK
                     KRIAEYKTLLKV"
     gene            complement(6151. .6900)
                     /gene="ylpA"
     CDS             complement(6151. .6900)
                     /gene="ylpA"
                     /codon_start=1
                     /transl_table=11
                     /product="lipoprotein YlpA"
                     /protein_id="AAD16806.1"
                     /db_xref="PID:g4324329"
                     /db_xref="GI:4324329"
                     /translation="MEDDMKKNMKLIAITAVLSSVLVLSGCGAMSTAIKKRNLEVKTQ
                     MSETIWLEPSSQKTVYLQIKNTSDKNMLGLAPKITKAVQDKGYTVTSSPEDAHYWIQA
                     NVLKADKMDLREAEGFLSQGYQGAALGAALGAGITGYNSNSAGASLGVGLAAGLVGMV
                     ADAMVEDINYTMVTDVQISEKTDTPLQTDNVAALKQGTSGYKVQTSTQTGNKHQYQTR
                     VVSSANKVNLKFEEAQPVLEDQLAKSIANIL"
     gene            complement(7370. .7918)
                     /gene="yopQ"
     CDS             complement(7370. .7918)
                     /gene="yopQ"
                     /codon_start=1
                     /transl_table=11
                     /product="YopQ"
                     /protein_id="AAD16807.1"
                     /db_xref="PID:g4324330"
                     /db_xref="GI:4324330"
                     /translation="MFIKDAYNMRALCTALEQSAPDTIINTSKEENNSYYCATAHLLR
                     TDVCSLVNRVGIEPLKSGSILSTLEELWQAVGIIYRLYEWQHVSDIDTNFKKLPNNSD
                     FGLVFSVLDCDIGYVITGKKDSKGNIELYDPKNSLLIENDDIKKYLYDENFHRFCIML
                     IISKSELEELSRESCDQKCIMG"
     gene            8439. .9407

alignment_scores:
           Quality:  286.00            Length:    214
             Ratio:  1.907               Gaps:      9
 Percent Similarity:  70.093   Percent Identity: 38.318

alignment_block:
 US-09-068-804-14 x AF102990   ..

 Align seg 1/1  to: AF102990  from: 1  to: 69673

       1 MetArgAspCysLeuAsnAsnGlyAsnProValLeuAsnValGlyAlaSe 17
         :::||||||||||||||:::      :::      |||:::::::::::::
   13389 TTACGCGATTGCCTGGACCGACAAGCCCATGAGCTAGAACTAAATAATCT 13438

      17 rGlyLeuThrThrLeuProAspArgLeuProProHisIleThrThrLeuV 34
          ||||||:::::||||||:::   ||||||||||||:::   :::||||
   13439 GGGGCTGAGTTCTTTGCCGGAA...TTACCTCCGCATTTAGAGAGTTTAG 13485

      34 alIleProAspAsnAsnLeuThrSerLeuProGluLeuProGluGlyLeu 50
         ||   :::  |||:::||||||   |||||||||||||||:::::|||
   13486 TGGCGTCATGTAATTCTCTTACAGAATTACCGGAATTGCCGCAGAGCCTG 13535

      51 ArgGluLeuGluValSerGlyAsnLeuGlnLeuThrSerLeuProSerLe 67
         :::   |||:::|||:::::|||   :::|||:::::|||:::::||
   13536 AAATCACTTCAAGTTGATAATAAC...AATCTGAAGGCATTATCCGATTT 13582

      67 uProGlnGlyLeuGlnLysLeuTrpAlaTyrAsnAsnTrpLeuAlaSerL 84
         ||||:::   |||:::   |||   |||   ||||||   |||:::   |
   13583 ACCTCCTTTACTGGAATATTTAGGTGCCGCTAATAATCAGCTGGAAGAAT 13632

      84 euProThrLeuProProGly......LeuGlyAspLeuAlaValSerAsn 98
         |||||   |||:::   :::      |||   ::::::::::|||:::|||
   13633 TACCAGAGTTGCAAAACTCGTCCTTCTTGACATCTATTGATGTTGATAAC 13682
```

```
      99 AsnGlnLeuThrSerLeuProGluMetProProAlaLeuArgGluLeuAr 115
         |||   |||::::::|||||||::::::|||||||::|||   |||:::
   13683 AATTCACTGAAAACATTACCTGATTTACCTCCTTCACTG...GAATTTCT 13729

     115 gValSerGlyAsnAsn.............LeuThrSerLeuArgAlaLeuP 128
          ::::::||||||||              |||:::   |||::::::||||
   13730 TGCTGCTGGTAATAATCAGCTGGAAGAATTGTCAGAGTTGCAAAACTTGC 13779

     128 roSerGlyLeuGlnLysLeuTrpAlaTyrAsnAsnArgLeuThrSerLeu 144
         ||      |||       :::   |||   ||||||   |||:::::|||
   13780 CCTTC...TTGACTGCGATTTATGCTGATAACAATTCACTGAAAACATTA 13826

     145 ProGluMetSerProGlyLeuGlnGluLeuAspValSerHisAsn..... 159
         |||::::::::::|||:::|||:::   |||:::|||   :::|||
   13827 CCCGATTTACCCCCTTCCCTGAAAACACTTAATGTCAGAGAAAATTATTT 13876

     160 ....GlnLeuThrArgLeuProGlnSerLeuThrGlyLeuSerSerAlaA 175
          :::|||:::   ||||||||||||||||||   |||:::   ::::
   13877 AACTGATCTGCCAGAATTACCGCAGAGTTTAACCTTCTTAGATGTTTCTG 13926

     175 laArgValTyr.................................Leu 179
         ::   ::::::                                 |||
   13927 ATAATATTTTTTCTGGATTATCGGAATTGCCACCAAACTTGTATAATCTC 13976

     180 AspGlyAsnProLeuSerValArgThrArgAspArgLeuCys 193
         ::::::::::::      :::|||:::         ||||||
   13977 AATGCATCCAGCAATGAAATAAGATCT.........TTATGC 14009
```

seq_name: gb_ba2:AF053946

seq_documentation_block:

```
LOCUS       AF053946    70504 bp    DNA   circular  BCT       06-OCT-1998
DEFINITION  Yersinia pestis plasmid pCD1, complete plasmid sequence.
ACCESSION   AF053946
NID         g2996222
VERSION     AF053946.1  GI:2996222
KEYWORDS    .
SOURCE      Yersinia pestis.
  ORGANISM  Yersinia pestis
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Yersinia.
REFERENCE   1  (bases 1 to 70504)
  AUTHORS   Hu,P., Elliott,J., McCready,P., Skowronski,E., Garnes,J.,
            Kobayashi,A., Brubaker,R.R. and Garcia,E.
  TITLE     Structural organization of virulence-associated plasmids of
            Yersinia pestis
  JOURNAL   J. Bacteriol. 180 (19), 5192-5202 (1998)
  MEDLINE   98422474
REFERENCE   2  (bases 1 to 70504)
  AUTHORS   Hu,P., Elliott,J., McCready,P., Skowronski,E., Garnes,J.,
            Kobayashi,A., Carrano,A.V., Brubaker,R. and Garcia,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAR-1998) Biology and Biotechnology, Lawrence
            Livermore National Lab, L452, 7000 East Ave., Livermore, CA 94550,
            USA
FEATURES             Location/Qualifiers
     source          1. .70504
                     /organism="Yersinia pestis"
                     /plasmid="pCD1"
                     /strain="KIM"
                     /db_xref="taxon:632"
     CDS             87. .1109
                     /codon_start=1
                     /transl_table=11
                     /product="transposase"
                     /protein_id="AAC62546.1"
                     /db_xref="PID:g2996223"
                     /db_xref="GI:2996223"
                     /translation="MVTFETVMEIKILHKQGMSSRAIARELGISRNTVKRYLQAKSEP
                     PKYTPRPAVASLLDEYRDYIRQRIADAHPYKIPATVIAREIRDQGYRGGMTILRAFIR
                     SLSVPQEQEPAVRFETEPGRQMQVDWGTMRNGRSPLHVFVAVLGYSRMLYIEFTDNMR
                     YDTLETCHRNAFRFFGGVPREVLYDNMKTVVLQRDAYQTGQHRFHPSLWQFGKEMGFS
                     PRLCRPFRAQTKGKVERMVQYTRNSFYIPLMTRLRPMGITVDVETANRHGLRWLHDVA
                     NQRKHETIQARPCDRWLEEQQSMLALPPEKKEYDVHLDENLVNFDKHPLHHPLSIYDS
                     FCRGVA"
     CDS             1109. .1888
                     /codon_start=1
                     /transl_table=11
                     /product="transposase"
                     /protein_id="AAC62557.1"
                     /db_xref="PID:g2996234"
                     /db_xref="GI:2996234"
                     /translation="MMELQHQRLMALAGQLQLESLISAAPALSQQAVDQEWSYMDFLE
                     HLLHEEKLARHQRKQAMYTRMAAFPAVKTFEEYDFTFATGAPQKQLQSLRSLSFIERN
                     ENIVLLGPSGVGKTHLAIAMGYEAVRAGIKVRFTTAADLLLQLSTAQRQGRYKTTLQR
                     GVMAPRLLIIDEIGYLPFSQEEAKLFFQVIAKRYEKSAMILTSNLPFGQWDQTFAGDA
                     ALTSAMLDRILHHSHVVQIKGESYRLRQKRKAGVIAEANPE"
     CDS             complement(1939. .2343)
                     /codon_start=1
                     /transl_table=11
                     /product="transposase"
                     /protein_id="AAC62568.1"
                     /db_xref="PID:g2996245"
                     /db_xref="GI:2996245"
                     /translation="MTEYQASERRGCRIMGISRSLLHYCPNTARDIPVVEVLQKLAHQ
                     YPAYGFGLMFNKLRQSGLPWNVKRVYRVYRLLKLNFRRKGKKRLPNRHPQPLAIPLKM
                     NHCWSVDFMSDALTDGRRFRLFNVVEIFVNDG"
     gene            complement(2379. .2645)
                     /gene="lcrS"
     CDS             complement(2379. .2645)
                     /gene="lcrS"
                     /codon_start=1
                     /transl_table=11
                     /product="low calcium response locus protein S homolog"
                     /protein_id="AAC62579.1"
                     /db_xref="PID:g2996256"
                     /db_xref="GI:2996256"
                     /translation="MKKARFTETQILRVLKEVEGGRHVKDVCRENGGSEASYYNWKSK
                     YGGMESSDIKRMKEREEENRRLKQMYASLSLDHEILKDVVAKKL"
     gene            complement(3193. .3540)
                     /gene="yscM"
     CDS             complement(3193. .3540)
                     /gene="yscM"
                     /codon_start=1
                     /transl_table=11
                     /product="Yop proteins translocation protein M homolog"
                     /protein_id="AAC62586.1"
                     /db_xref="PID:g2996263"
                     /db_xref="GI:2996263"
                     /translation="MKINTLQSLINQQITQVGHGGQAGRLTETNPLTENSHQISTAEK
                     AFANEVLEHVKNTALSRHDIACLLPRVSNLELKQGKAGEVIVTGLRTEQLSLSDAKLL
                     LEAAMRQDTAADG"
     gene            complement(3765. .4430)
                     /gene="yscL"
     CDS             complement(3765. .4430)
                     /gene="yscL"
                     /codon_start=1
                     /transl_table=11
                     /product="Yop proteins translocation protein L  homolog"
                     /protein_id="AAC62596.1"
                     /db_xref="PID:g2996273"
                     /db_xref="GI:2996273"
                     /translation="MSQTCQTGYAYMQPFVQIIPSNLSLACGLRILRAEDYQSSLTTE
                     ELISAAKQDAEKILADAQEVYEQQKQLGWQAGMDEARTLQATLIHETQLQCQQFYRHV
                     EQQMSEVVLLAVRKILNDYDQVAMTLQVVREALALVSNQKQVVVRVNPDQAGAIREQI
                     AKVHKDFPEISYLEVTADARLDQGGCILETEVGIIDASIDGQIEALSRAISTTLGQMK
                     VTE"
     gene            complement(4376. .5005)
                     /gene="yscK"
     CDS             complement(4376. .5005)
                     /gene="yscK"
                     /codon_start=1
                     /transl_table=11
                     /product="Yop proteins translocation protein K homolog"
                     /protein_id="AAC62605.1"
```

```
                /db_xref="PID:g2996282"
                /db_xref="GI:2996282"
                /translation="MMENYITSFQLRFCPAAYLHLEQLPSLWRSILPYLPQWRDSAHL
                NAALLDEFSLDTDYEEPHGLGALPLQPQSQLELLLCRLGLVLHGEAIRRCVLASPLQQ
                LLTLVNQETLRQIIVQHELLIGPWPTHWQRPLPTEIESRTMIQSGLAFWLAAMEPQPQ
                AWCKRLSLRLPLATPSEPWLVAESQRPLAQTLCHKLVKQVTPTCSHLFK"
gene            complement(5005. .5739)
                /gene="yscJ"
CDS             complement(5005. .5739)
                /gene="yscJ"
                /codon_start=1
                /transl_table=11
                /product="Yop proteins translocation lipoprotein J"
                /protein_id="AAC62607.1"
                /db_xref="PID:g2996284"
                /db_xref="GI:2996284"
                /translation="MKVKTSLSTLILILFLTGCKVDLYTGISQKEGNEMLALLRQEGL
                SADKEPDKDGKIKLLVEESDVAQAIDILKRKGYPHESFSTLQDVFPKDGLISSPIEEL
                ARLNYAKAQEISRTLSEIDGVLVARVHVVLPEEQNNKGKKGVAASASVFIKHAADIQF
                DTYIPQIKQLVNNSIEGLAYDRISVILVPSVDVRQSSHLPRNTSILSIQVSEESKGHL
                IGLLSLLILLLPVTNLAQYFWLQRKK"
gene            complement(5746. .6093)
                /gene="yscI"
CDS             complement(5746. .6093)
                /gene="yscI"
                /codon_start=1
                /transl_table=11
                /product="Yop proteins translocation protein I homolog"
                /protein_id="AAC62608.1"
                /db_xref="PID:g2996285"
                /db_xref="GI:2996285"
                /translation="MPNIEIAQADEVIITTLEELGPAEPTTDQIMRFDAAMSEDTQGL
                GHSLLKEVSDIQKSFKTVKSDLHTKLAVSVDNPNDLMLMQWSLIRITIQEELIAKTAG
                RMSQNVETLSKGG"
gene            complement(6094. .6591)
                /gene="yscH"
CDS             complement(6094. .6591)
                /gene="yscH"
                /codon_start=1
                /transl_table=11
                /product="Yop proteins translocation protein H homolog"
                /protein_id="AAC62547.1"
                /db_xref="PID:g2996224"
                /db_xref="GI:2996224"
                /translation="MTVTLNRGSITSLMSSSQAVSTLQPVASELKTQLENKLKSESAE
                KIREVLWQQYYASNPPDHAVLEVLATPVREALLARFGQHQGSVVPAIDLPELRSVLQQ
                FDSFGKRWEAILLQVLEGIKPNESQVGLPYLSELINKELMILLPSNSIVDSLLHNSHQ
                IDMDT"
gene            complement(6588. .6935)
                /gene="yscG"
CDS             complement(6588. .6935)
                /gene="yscG"
                /codon_start=1
                /transl_table=11
                /product="Yop proteins translocation protein G homolog"
                /protein_id="AAC62548.1"
                /db_xref="PID:g2996225"
                /db_xref="GI:2996225"
                /translation="MKYKLNVLLAEIALIGTGNHYHEEANCIAEWLHLKGEEEAVQLI
                RLSSLMNRGDYASALQQGNKLAYPDLEPWLALCEYRLGLGSALESRLNRLARSQDPRI
                QTFVNGMREQLKT"
gene            complement(6937. .7200)
                /gene="yscF"
CDS             complement(6937. .7200)
                /gene="yscF"
                /codon_start=1
                /transl_table=11
                /product="Yop proteins translocation protein F homolog"
                /protein_id="AAC62549.1"
                /db_xref="PID:g2996226"
                /db_xref="GI:2996226"
                /translation="MSNFSGFTKGTDIADLDAVAQTLKKPADDANKAVNDSIAALKDK
                PDNPALLADLQHSINKWSVIYNINSTIVRSMKDLMQGILQKFP"
gene            complement(7201. .7401)
                /gene="yscE"
CDS             complement(7201. .7401)
                /gene="yscE"
                /codon_start=1
                /transl_table=11
                /product="Yop proteins translocation protein E homolog"
                /protein_id="AAC62550.1"
                /db_xref="PID:g2996227"
                /db_xref="GI:2996227"
                /translation="MTQLEEQLHNVETVRSITMQLEMALTKLKKDMMRGGDAKQYQVW
                QRESKALESAIAIIHYVAGDLK"
gene            complement(7398. .8657)

alignment_scores:
          Quality:  266.00          Length:    254
            Ratio:  1.762             Gaps:     10
Percent Similarity:  59.449  Percent Identity: 32.677

alignment_block:
US-09-068-804-14 x AF053946   ..

Align seg 1/1  to: AF053946  from: 1  to: 70504

      1 MetArgAspCysLeuAsnAsnGlyAsnProValLeuAsnValGlyAlaSe 17
        :::|||||||||||||:::       :::       |||:::::::::::::
  31065 TTACGAGATTGCCTGGACCGACAAGCCCATGAGCTAGAACTAAATAATCT 31114

     17 rGlyLeuThrThrLeuProAspArgLeuProProHisIleThrThrLeuV 34
        ||||||:::::|||||||:::   |||||||||||||:::   :::||||
  31115 GGGGCTGAGTTCTTTGCCGGAA...TTACCTCCGCATTTAGAGAGTTTAG 31161

     34 alIleProAspAsnAsnLeuThrSerLeuProGluLeuProGluGlyLeu 50
        ||   :::   |||:::||||||   |||||||||||||||:::::|||
  31162 TGGCGTCATGTAATTCTCTTACAGAATTACCGGAATTACCGCAGAGCCTG 31211

     51 ArgGluLeuGluValSerGlyAsnLeuGlnLeuThrSerLeuProSerLe 67
        :::   |||   |||:::::::|||   :::|||:::::::|||:::::||
  31212 AAATCACTTCTAGTTGATAATAAC...AATCTGAAGGCATTATCCGATTT 31258

     67 uPro.............................GlnGlyLeuGlnLysL 74
        ||||                           :::   |||:::||||
  31259 ACCACCTTTACTGGAATATTTAGGTGTCTCTAATAATCAGCTGGAAAAAT 31308

     74 eu..............TrpAlaTyr.....................Asn 78
        ||                 ::::::                   |||
  31309 TGCCAGAGTTGCAAAACTCGTCCTTCTTGAAAATTATTGATGTTGATAAC 31358

     79 AsnTrpLeuAlaSerLeuProThrLeuProProGlyLeuGlyAspLeuAl 95
        |||   |||     ||||||   ||||||||||:::|||      :::||
  31359 AATTCACTGAAAAAACTACCTGATTTACCTCCTTCACTGGAGTTTATTGC 31408

     95 aValSerAsnAsnGlnLeuThrSerLeuProGluMetPro......ProA 110
        |:::::|||||||||||      ||||||||:::::      |||
  31409 TGCTGGTAATAATCAGCTGGAAGAATTGCCAGAGTTGCAAAACTTGCCCT 31458

    110 laLeuArgGluLeuArgValSerGlyAsnAsnLeuThrSerLeuArgAla 126
          |||   ::::::  :::::::::|||:::|||:::   |||   :::
  31459 TCTTGACTGCGATTTATGCTGATAACAATTCACTGAAAAAACTACCTGAT 31508

    127 LeuProSerGlyLeuGln............................... 132
        ||||||   :::|||:::
  31509 TTACCTCTTTCACTGGAATCTATTGTTGCTGGTAATAATATTCTGGAAGA 31558

    133 ..............................LysLeuTrpAlaTyrA 138
                                     ::::::   |||   |
  31559 ATTGCCAGAGTTGCAAAACTTGCCCTTCTTGACTACGATTTATGCTGATA 31608

    138 snAsnArgLeuThrSerLeuProGluMetSerProGlyLeuGlnGluLeu 154
        |||||   |||:::::|||||||:::::::::|||:::|||:::::|||
  31609 ACAATTTACTGAAAACATTACCCGATTTACCCCCTTCCCTGGAAGCACTT 31658

    155 AspValSerHisAsn.........GlnLeuThrArgLeuProGlnSerLe 168
```

```
      :::|||   :::|||         :::|||:::   |||||||||||||||
  31659 AATGTCAGAGATAATTATTTAACTGATCTGCCAGAATTACCGCAGAGTTT 31708

    168 uThrGlyLeuSerSerAlaAlaArgVal..................... 177
        ||||   |||:::   ::::::   :::
  31709 AACCTTCTTAGATGTTTCTGAAAATATTTTTTCTGGATTATCGGAATTGC 31758

    178 .............TyrLeuAspGlyAsnProLeuSerValArgThrArg 189
                     ||||||:::::::::::::      :::|||:::
  31759 CACCAAACTTGTATTATCTCAATGCATCCAGCAATGAAATAAGATCC... 31805

    190 AspArgLeuCys 193
            ||||||
  31806 ......TTATGC 31811

seq_name: gb_ba2:AF074612

seq_documentation_block:
LOCUS       AF074612    70559 bp    DNA   circular  BCT       03-NOV-1998
DEFINITION  Yersinia pestis plasmid pCD1, complete plasmid sequence.
ACCESSION   AF074612
NID         g3822037
VERSION     AF074612.1  GI:3822037
KEYWORDS    .
SOURCE      Yersinia pestis.
  ORGANISM  Yersinia pestis
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Yersinia.
REFERENCE   1  (bases 1 to 70559)
  AUTHORS   Perry,R.D., Straley,S.C., Fetherston,J.D., Rose,D.J., Gregor,J. and
            Blattner,F.R.
  TITLE     DNA sequencing and analysis of the low-Ca2+-response plasmid pCD1
            of Yersinia pestis KIM5
  JOURNAL   Infect. Immun. 66 (10), 4611-4623 (1998)
  MEDLINE   98427122
REFERENCE   2  (bases 1 to 70559)
  AUTHORS   Perry,R.D., Straley,S.C., Fetherston,J.D., Rose,D.J., Gregor,J. and
            Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUN-1998) Microbiology and Immunology, University of
            Kentucky, MS415 Medical Center, Lexington, KY 40536-0084, USA
COMMENT     Unique identifiers, starting at Y0001, are assigned to each gene of
            pCD1.
FEATURES             Location/Qualifiers
     source          1. .70559
                     /organism="Yersinia pestis"
                     /plasmid="pCD1"
                     /strain="KIM5"
                     /db_xref="taxon:632"
     gene            57. .368
                     /gene="Y0001"
     CDS             57. .368
                     /gene="Y0001"
                     /note="o103; 43 pct identical (0 gaps) to 100 residues of
                     an approx. 200 aa protein GENPEPT: gi|537126, orf_o198
                     Escherichia coli"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAC69758.1"
                     /db_xref="PID:g3822038"
                     /db_xref="GI:3822038"
                     /translation="MHQQSRGAAGSRTLSLLMRQSGYNVVRWLARRLMRECGLASRQP
                     GKPRYRGEREVSLASPDLLKRQFKPSEPNRVWSGYISYIKVNGGWCYLALVIDLYSFH
                     W"
     gene            665. .1033
                     /gene="nuc"
                     /note="Y0002"
     CDS             665. .1033
                     /gene="nuc"
                     /codon_start=1
                     /transl_table=11
                     /product="endonuclease"
                     /protein_id="AAC69759.1"
                     /db_xref="PID:g3822039"
                     /db_xref="GI:3822039"
                     /translation="MDTKLLQHTPIGTMVDYRPVNTKSGGKRLRRCPDFVIHYRMDLL
                     VNAGIPVRTVNSFKALHDKVIIVDGKNTQMGSFNFSQAAVQSNSENVLIIWGDFTVVQ
                     AYLQYWQSRWNKGTDWRSSY"
     gene            1171. .1425
                     /gene="repB"
                     /note="Y0003"
     CDS             1171. .1425
                     /gene="repB"
                     /codon_start=1
                     /transl_table=11
                     /product="replication protein"
                     /protein_id="AAC69760.1"
                     /db_xref="PID:g3822040"
                     /db_xref="GI:3822040"
                     /translation="MSQIENAVTSSSKRAYRKGNPLTGAEKQRMSVSRKKETHKAINV
                     FIQNDLKNELLQLCEDSGLTQTEMIERWIQREKAARTNAA"
     misc_RNA        complement(1560. .1649)
                     /note="antisense RNA"
                     /product="copA"
     gene            1667. .1741
                     /gene="tap"
                     /note="Y0004"
     CDS             1667. .1741
                     /gene="tap"
                     /codon_start=1
                     /transl_table=11
                     /product="repA translation protein"
                     /protein_id="AAC69761.1"
                     /db_xref="PID:g3822041"
                     /db_xref="GI:3822041"
                     /translation="MFRKVQYLLRLLLLPCNISAGRCD"
     gene            1734. .2600
                     /gene="repA"
                     /note="Y0005"
     CDS             1734. .2600
                     /gene="repA"
                     /codon_start=1
                     /transl_table=11
                     /product="replication protein"
                     /protein_id="AAC69762.1"
                     /db_xref="PID:g3822042"
                     /db_xref="GI:3822042"
                     /translation="MTNHQALFTHHYRQVKNPNPEFTPREGKKTLPFCRKLMAKAEGF
                     TSRFDFSMHVAFARSLSLRHRMPPLLRRRAIDALLQGMCFHYDPLANRIQRSITNLAI
                     ECGLATESKSGNLSITRATRALRFLSELGLITYQTEYDPQIGCNIPTDITFTPALFSA
                     LDVSDVAVAAARRSRVEWENQQREKQRLPRLEMDELIAKAWRFVRERFRSYQTERKAH
                     GLKRARARRDVDRTRRDIEAIVNRQLTREIAEGRFVGNLDAVRREKARRVKERMLMSR
                     NNNYTRLATGAT"
     rep_origin      1741. .2935
                     /note="oriR"
     gene            complement(3427. .3645)
                     /gene="Y0006"
     CDS             complement(3427. .3645)
                     /gene="Y0006"
                     /note="Orf5 (f72); 42 pct identical (0 gaps) to 33
                     residues of an approx. 216 aa protein GENPEPT: gi|2055297,
                     lmp2 Xenopus laevis"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAC69763.1"
                     /db_xref="PID:g3822043"
                     /db_xref="GI:3822043"
                     /translation="MRSPLAYGCSVYTVNVVTQLHNDIHKSAYKHKRLYHVLTGQASS
                     AQPVFSMPHRGMKAPADINYNSVDNWRR"
     gene            4758. .5186
                     /gene="Y0008"
     CDS             4758. .5186
                     /gene="Y0008"
                     /note="Orf7 (o142); 31 pct identical (1 gap) to 48
```

```
                residues of an approx. 104 aa protein GENPEPT: gi|2149940,
                orf1 P. syringae"
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AAC69764.1"
                /db_xref="PID:g3822044"
                /db_xref="GI:3822044"
                /translation="MINTTFTELLPKIASHFGLDKLSQDEYGLCELILNDRVVIMLRA
                DEILNRLTLLGPILGFSGPEARSAASQLFFCYSINALNKDGPCFAWSEELGLIAFKHL
                SLDELNVENVSKEIANFYDWLSLVSLPAETAPSYSIYSIG"
gene            5204. .7402
                /gene="ypkA"
                /note="Y0009"
CDS             5204. .7402
                /gene="ypkA"
                /codon_start=1
                /transl_table=11
                /product="secreted protein kinase"
                /protein_id="AAC69765.1"
                /db_xref="PID:g3822045"
                /db_xref="GI:3822045"
                /translation="MKSVKIMGTMPPSISLAKAHERISQHWQNPVGELNIGGKRYRII
                DNQVLRLNPHSGFSLFREGVGKIFSGKMFNFSIARNLTDTLHAAQKTTSQELRSDIPN
                ALSNLFGAKPQTELPLGWKGEPLSGAPDLEGMRVAETDKFAEGESHISIIETKDKQRL
                VAKIERSIAEGHLFAELEAYKHIYKTAGKHPNLANVHGMAVVPYGNRKEEALLMDEVD
                GWRCSDTLRTLADSWKQGKINSEAYWGTIKFIAHRLLDVTNHLAKAGVVHNDIKPGNV
                VFDRASGEPVVIDLGLHSRSGEQPKGFTESFKAPELGVGNLGASEKSDVFLVVSTLLH
                CIEGFEKNPEIKPNQGLRFITSEPAHVMDENGYPIHRPGIAGVETAYTRFITDILGVS
                ADSRPDSNEARLHEFLSDGTIDEESAKQILKDTLTGEMSPLSTDVRRITPKKLRELSD
                LLRTHLSSAATKQLDMGGVLSDLDTMLVALDKAEREGGVDKDQLKSFNSLILKTYRVI
                EDYVKGREGDTKNSSTEVSPYHRSNFMLSIVEPSLQRIQKHLDQTHSFSDIGSLVRAH
                KHLETLLEVLVTLSQQGQPVSSETYGFLNRLTEAKITLSQQLNTLQQQQESAKAQLSI
                LINRSGSWADVARQSLQRFDSTQPVVKFGTEQYTAIHRQMMAAHAAITLQEVSEFTDD
                MRNFTVDSIPLLIQLGRSSLMDEHLVEQREKLRELTTIAERLNRLEREWM"
gene            7798. .8664
                /gene="yopJ"
                /note="Y0010"
CDS             7798. .8664
                /gene="yopJ"
                /codon_start=1
                /transl_table=11
                /product="targeted effector protein"
                /protein_id="AAC69766.1"
                /db_xref="PID:g3822046"
                /db_xref="GI:3822046"
                /translation="MIGPISQINISGGLSEKETSSLISNEELKNIITQLETDISDGSW
                FHKNYSRMDVEVMPALVIQANNKYPEMNLNLVTSPLDLSIEIKNVIENGVRSSRFIIN
                MGEGGIHFSVIDYKHINGKTSLILFEPANFNSMGPAMLAIRTKTAIERYQLPDCHFSM
                VEMDIQRSSSECGIFSLALAKKLYIERDSLLKIHEDNIKGILSDGENPLPHDKLDPYL
                PVTFYKHTQGKKRLNEYLNTNPQGVGTVVNKKNETIVNRFDNNKSIVDGKELSVSVHK
                KRIAEYKTLLKV"
gene            9858. .10064
                /gene="Y0012"
CDS             9858. .10064
                /gene="Y0012"
                /note="o68; 45 pct identical (0 gaps) to 68 residues of an
                approx. 560 aa protein GENPEPT: gi|2182435, hypothetical"
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AAC69767.1"
                /db_xref="PID:g3822047"
                /db_xref="GI:3822047"
                /translation="MAVGRKNWLFAGSLRAGQRMASILSLLETAKLNGHDPYVWLRDV
                LTRLPTWPNSQLNALLPYAENRFS"
gene            10347. .11753
                /gene="yopH"
                /note="Y0013"
CDS             10347. .11753
                /gene="yopH"
                /codon_start=1
```

```
alignment_scores:
            Quality:  266.00              Length:    254
              Ratio:  1.762                 Gaps:     10
 Percent Similarity:  59.449   Percent Identity: 32.677

alignment_block:
 US-09-068-804-14 x AF074612   ..

 Align seg 1/1  to: AF074612  from: 1  to: 70559

      1 MetArgAspCysLeuAsnAsnGlyAsnProValLeuAsnValGlyAlaSe 17
        :::|||||||||||||:::      :::      |||:::::::::::::
  43673 TTACGAGATTGCCTGGACCGACAAGCCCATGAGCTAGAACTAAATAATCT 43722

     17 rGlyLeuThrThrLeuProAspArgLeuProProHisIleThrThrLeuV 34
        ||||||::::::|||||:::   |||||||||||||:::   :::||||
  43723 GGGGCTGAGTTCTTTGCCGGAA...TTACCTCCGCATTTAGAGAGTTTAG 43769

     34 alIleProAspAsnAsnLeuThrSerLeuProGluLeuProGluGlyLeu 50
        ||   :::   |||:::|||||   |||||||||||||||:::::|||
  43770 TGGCGTCATGTAATTCTCTTACAGAATTACCGGAATTACCGCAGAGCCTG 43819

     51 ArgGluLeuGluValSerGlyAsnLeuGlnLeuThrSerLeuProSerLe 67
        :::   |||   |||::::::|||   :::|||::::::|||::::::||
  43820 AAATCACTTCTAGTTGATAATAAC...AATCTGAAGGCATTATCCGATTT 43866

     67 uPro.............................GlnGlyLeuGlnLysL 74
        ||||                              :::   |||:::||||
  43867 ACCACCTTTACTGGAATATTTAGGTGTCTCTAATAATCAGCTGGAAAAAT 43916

     74 eu..............TrpAlaTyr.....................Asn 78
        ||                  ::::::                   |||
  43917 TGCCAGAGTTGCAAAACTCGTCCTTCTTGAAAATTATTGATGTTGATAAC 43966

     79 AsnTrpLeuAlaSerLeuProThrLeuProProGlyLeuGlyAspLeuAl 95
        |||   |||      ||||||   ||||||||||:::|||      :::||
  43967 AATTCACTGAAAAAACTACCTGATTTACCTCCTTCACTGGAGTTTATTGC 44016

     95 aValSerAsnAsnGlnLeuThrSerLeuProGluMetPro......ProA 110
        |::::::||||||||||||      ||||||||||::::::      |||
  44017 TGCTGGTAATAATCAGCTGGAAGAATTGCCAGAGTTGCAAAACTTGCCCT 44066

    110 laLeuArgGluLeuArgValSerGlyAsnAsnLeuThrSerLeuArgAla 126
          |||   ::::::   ::::::::::|||:::|||:::   |||   :::
  44067 TCTTGACTGCGATTTATGCTGATAACAATTCACTGAAAAAACTACCTGAT 44116

    127 LeuProSerGlyLeuGln.............................. 132
        ||||||   :::|||:::
  44117 TTACCTCTTTCACTGGAATCTATTGTTGCTGGTAATAATATTCTGGAAGA 44166

    133 ..............................LysLeuTrpAlaTyrA 138
                                      ::::::   |||   |
  44167 ATTGCCAGAGTTGCAAAACTTGCCCTTCTTGACTACGATTTATGCTGATA 44216

    138 snAsnArgLeuThrSerLeuProGluMetSerProGlyLeuGlnGluLeu 154
        |||||   |||::::::|||||:::::::::|||:::|||::::::|||
  44217 ACAATTTACTGAAAACATTACCCGATTTACCCCCTTCCCTGGAAGCACTT 44266

    155 AspValSerHisAsn.........GlnLeuThrArgLeuProGlnSerLe 168
        :::|||   :::|||          :::|||:::   ||||||||||||||
  44267 AATGTCAGAGATAATTATTTAACTGATCTGCCAGAATTACCGCAGAGTTT 44316

    168 uThrGlyLeuSerSerAlaAlaArgVal..................... 177
        ||||   |||:::   ::::::   :::
  44317 AACCTTCTTAGATGTTTCTGAAAATATTTTTTCTGGATTATCGGAATTGC 44366

    178 .............TyrLeuAspGlyAsnProLeuSerValArgThrArg 189
                     ||||||:::::::::::::      :::|||:::
  44367 CACCAAACTTGTATTATCTCAATGCATCCAGCAATGAAATAAGATCC... 44413

    190 AspArgLeuCys 193
            ||||||
  44414 ......TTATGC 44419
```

```
seq_name: gb_pr1:AB016816

seq_documentation_block:
LOCUS       AB016816     3159 bp    mRNA            PRI       20-FEB-1999
DEFINITION  Homo sapiens MASL1 mRNA, complete cds.
ACCESSION   AB016816
NID         g4239894
VERSION     AB016816.1  GI:4239894
KEYWORDS    MASL1.
SOURCE      Homo sapiens tissue_lib:human normal fetal-brain cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3159)
  AUTHORS   Inazawa,J., Nakamura,Y. and Sakabe,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-AUG-1998) to the DDBJ/EMBL/GenBank databases. Johji
            Inazawa, Institute of Medical Science, University of Tokyo,
            Laboratory of Genome Medicine, Human Genome Center; Minato-ku,
            Shirokanedai 4-6-1, Tokyo, Tokyo 108-8639, Japan
            (E-mail:johinaz@ims.u-tokyo.ac.jp, Tel:81-3-5449-5237,
            Fax:81-3-5449-5406)
REFERENCE   2  (bases 1 to 3159)
  AUTHORS   Inazawa,J., Nakamura,Y. and Sakabe,T.
  TITLE     Identification of a novel gene, MASL1, within an amplicon at 8p23.1
            in malignant fibrous histiocytoma
  JOURNAL   Published Only in DataBase (1999) In press
FEATURES             Location/Qualifiers
     source          1. .3159
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8p23.1"
                     /tissue_lib="human normal fetal-brain"
     gene            1. .3159
                     /gene="MASL1"
     CDS             1. .3159
                     /gene="MASL1"
                     /codon_start=1
                     /protein_id="BAA74737.1"
                     /db_xref="PID:d1038470"
                     /db_xref="PID:g4239895"
                     /db_xref="GI:4239895"
                     /translation="MAGMDSGNLKTARLWRDAALRARKLRSNLRQLTLTAAGACPGAG
                     ADALESPASPQLVLPANLGDIEALNLGNNGLEEVPEGLGSALGSLRVLVLRRNRFARL
                     PPAVAELGHHLTELDVSHNRLTALGAEVVSALRELRKLNLSHNQLPALPAQLGALAHL
                     EELDVSFNRLAHLPDSLSCLSRLRTLDVDHNQLTAFPRQLLQLVALEELDVSSNRLRG
                     LPEDISALRALKILWLSGAELGTLPAGFCELASLESLMLDNNGLQALPAQFSCLQRLK
                     MLNLSSNLFEEFPAALLPLAGLEELYLSRNQLTSVPSLISGLGRLLTLWLDNNRIRYL
                     PDSIVELTGLEELVLQGNQIAVLPDHFGQLSRVGLWKIKDNPLIQPPYEVCMKGIPYI
                     AAYQKELAHSQPAVQPRLKLLLMGHKAAGKTLLRHCLTEERVEGCPGGGDKEKCYPPS
                     PPPVSKGIEVTSWTADASRGLRFIVYDLAGDESYEVIQPFFLSPGALYVLVVNLATYE
                     PRHFPTTVGSFLHRVGARVPNAVVCIVGTHADLCGERELEEKCLDIHRQIALQEKHDA
                     EGLSRLAKVVDEALARDFELRSASPHAAYYGVSDKNLRRRKAHFQYLLNHRLQILSPV
                     LPVSCRDPRHLRRLRDKLLSVAEHREIFPNLHRVLPRSWQVLEELHFQPPQAQRLWLS
                     WWDSARLGLQAGLTEDRLQSALSYLHESGKLLYFEDSPALKEHVFHNLTRLIDILNVF
                     FQRDPSLLLHKLLLGTSGEGKAEGESSPPMARSTPSQELLRATQLHQYVEGFLLHGLL
                     PAHVIRLLLKPHVQAQQDLQLLLELLEKMGLCYCLNKPKGKPLNGSTAWYKFPCYVQN
                     EVPHAEAWINGTNLAGQSFVAEQLQIEYSFPFTFPPGLFARYSVQINSHVVHRSDGKF
                     QIFAYRGKVPVVVSYRPARGVLQPDTLSIASHASLPNIWTAWQAITPLVEELNVLLQE
                     WPGLHYTVHILCSKCLKRGSPNPHAFPGELLSQPRPEGVAEIICPKNGSERVNVALVY
                     PPTPTVISPCSKKNVGEKHRNQ"
BASE COUNT      596 a   1001 c    944 g    618 t
ORIGIN


alignment_scores:
            Quality:  170.50            Length:    281
              Ratio:  1.176               Gaps:     15
 Percent Similarity:  51.601   Percent Identity: 31.317

alignment_block:
```

```
US-09-068-804-14 x AB016816   ..

Align seg 1/1  to: AB016816  from: 1  to: 3159

     12 LeuAsnValGlyAlaSerGlyLeuThrThrLeuProAspArgLeuPro.. 27
        |||:::|||:::   :::   |||:::   ||||||||   |||:::
    487 CTGGATGTCAGCTTTAACCGGCTGGCGCACCTGCCTGACTCCCTCTCCTG 536

     28 ....ProHisIleThrThrLeuValIleProAspAsnAsnLeuThrSerL 43
            :::::::::   ||||||   :::   :::|||:::|||||||:::
    537 CCTCTCCCGCCTGCGCACCCTGGACGTGGATCACAACCAGCTCACTGCCT 586

     43 euPro...GluLeuProGlu......GlyLeuArgGluLeuGluValSer 56
        ::|||   :::|||   :::      :::|||   ||||||:::||||||
    587 TCCCCCGGCAGCTGCTGCAGCTGGTGGCCCTGGAGGAGCTGGACGTGTCC 636

     57 GlyAsnLeuGlnLeuThrSerLeuProSerLeuProGlnGlyLeuGlnLy 73
        :::|||                |||   :::||||||:::::::::
    637 AGCAAC.............CGGCTGCGGGGCCTGCCTGAGGATATCAGTGC 674

     73 sLeuTrpAlaTyrAsnAsn...TrpLeu......AlaSerLeuProThrL 87
         |||:::|||   :::      ||||||      |||   |||   ||||
    675 CCTGCGTGCCCTCAAGATCCTCTGGCTGAGTGGGGCCGAGCTTGGCACGC 724

     87 euProProGly.................LeuGlyAspLeuAlaValSer 97
        |||||:::|||                |||   :::|||   ::::::
    725 TGCCCGCCGGCTTCTGCGAGCTGGCCAGTTTGGAGAGCCTCATGCTAGAC 774

     98 AsnAsnGlnLeuThrSerLeuPro..........................  105
        ||||||   |||   :::||||||
    775 AACAACGGGCTGCAGGCTCTGCCCGCCCAGTTCAGCTGCCTGCAGCGGCT 824

    106 ................................GluMetProProAla.  110
                                       |||   |||:::|||
    825 CAAAATGCTCAACCTGTCCTCCAACCTCTTCGAGGAGTTCCCTGCCGCGC 874

    111 .................LeuArgGluLeuArgValSerGlyAsnAsnLeu 121
                         |||   ||||||   :::|||   |||:::|||
    875 TGCTGCCCCTGGCTGGTCTGGAGGAGCTCTACCTTAGTCGCAACCAGCTC 924

    122 ThrSerLeuArgAlaLeuProSerGlyLeuGlnLys.........LeuTr 135
        |||||:::   :::|||   |||||||||   :::         |||||
    925 ACCTCGGTGCCATCCCTTATCTCGGGCCTGGGCCGGCTTCTCACCTTGTG 974

    135 pAlaTyrAsnAsnArgLeuThrSerLeuProGluMetSer.........P 149
        |      |||||||||:::      ||||||:::              :
    975 GCTGGATAATAACCGCATCCGCTACCTGCCGGACTCCATCGTGGAGCTGA 1024

    149 roGlyLeuGlnGluLeuAspValSerHisAsnGlnLeuThrArgLeuPro 165
        ::||||||:::||||||   :::      ||||||:::::   ||||||
   1025 CCGGCCTGGAGGAGCTCGTGCTGCAGGGGAACCAGATCGCGGTGCTGCCC 1074

    166 GlnSerLeuThrGlyLeuSerSerAlaAla.................... 175
        :::   :::      ||||||   ::::::
   1075 GACCACTTTGGCCAGCTCTCCCGGGTGGGTTTGTGGAAGATCAAAGACAA 1124

    176 .....................ArgValTyrLeuAsp.GlyAsnProLeu 184
                              |||:::::   |||:::|||
   1125 CCCACTGATCCAGCCCCCCTACGAGGTCTGCATGAAGGGGATCCCCTACA 1174

    185 SerValArgThrArg.................AspArgLeuCysGlyHi 195
        |||      ||||||                :::|||   |||:::
   1175 TCGCAGCCTACCAGAAGGAACTGGCTCATTCCCAGCCGGCGGTGCAGCCC 1224

    195 sHis..............TrpProPheArgHisGlnAspThrLeuArgT 207
                         |||   :::|||   |||:::   |||   :
   1225 CGGCTCAAGCTGCTCCTGATGGGGCATAAGGCTGCAGGAAAGACTTTGCT 1274

    207 yrGlyGlyAlaPheArgProArgGluAlaArgAlaLeuHis 220
        :::::   |||         ||||||   |||:::   :::
   1275 GCGCCACTGCCTCACCGAGGAGAGAGTGGAGGGATGCCCAG 1315

seq_name: gb_ro:MMPLGPV
```

```
seq_documentation_block:
LOCUS       MMPLGPV      3584 bp    DNA             ROD       16-MAY-1997
DEFINITION  M.musculus platelet glycoprotein V gene.
ACCESSION   Z69595
NID         g2104844
VERSION     Z69595.1  GI:2104844
KEYWORDS    cis-acting element; platelet glycoprotein V; repeat region.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 3584)
  AUTHORS   Ravanat,C., Morales,M., Azorsa,D.O., Moog,S., Schuhler,S.,
            Grunert,P., Cazenave,J. and Lanza,F.
  TITLE     Rodent platelet glycoprotein V: gene structure and characterization
            of the protein
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3584)
  AUTHORS   Lanza,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-FEB-1996) Lanza F., Etablissement de Transfusion
            Sanguine de Strasbourg, INSERM U.311, 10, rue Spielmann,
            Strasbourg, FRANCE, 67065
REFERENCE   3  (bases 1 to 3584)
  AUTHORS   Ravanat,C., Morales,M., Azorsa,D.O., Moog,S., Schuhler,S.,
            Grunert,P., Loew,D., Van Dorsselaer,A., Cazenave,J.P. and Lanza,F.
  TITLE     Gene cloning of rat and mouse platelet glycoprotein V:
            identification of megakaryocyte-specific promoters and
            demonstration of functional thrombin cleavage
  JOURNAL   Blood 89 (9), 3253-3262 (1997)
  MEDLINE   97275136
FEATURES             Location/Qualifiers
     source          1. .3584
                     /organism="Mus musculus"
                     /strain="C57 Black/6"
                     /db_xref="taxon:10090"
                     /sex="Female"
                     /clone="lambda MV1"
                     /dev_stage="adult"
                     /tissue_type="liver"
                     /clone_lib="lambda DASH II genomic library"
     protein_bind    370. .375
                     /note="conserved in human and rat genes"
                     /citation=[1]
                     /function="cis-acting element"
                     /bound_moiety="GATA-1"
     protein_bind    complement(408. .413)
                     /note="conserved in human and rat genes"
                     /citation=[1]
                     /function="cis-acting element"
                     /bound_moiety="GATA-1"
     protein_bind    415. .422
                     /note="conserved in human and rat genes"
                     /citation=[1]
                     /function="cis-acting element"
                     /bound_moiety="Ets-1"
     protein_bind    complement(437. .444)
                     /note="conserved in human and rat genes"
                     /citation=[1]
                     /function="cis-acting element"
                     /bound_moiety="Ets-1"
     prim_transcript 482. .3584
                     /standard_name="platelet GPV"
                     /citation=[1]
     exon            483. .511
                     /citation=[1]
                     /number=1
     intron          512. .1405
                     /citation=[1]
                     /number=1
     exon            1406. .3584
                     /citation=[1]
                     /number=2
     sig_peptide     1408. .1455
     CDS             1408. .3111
                     /standard_name="platelet GPV"
                     /note="precursor"
                     /citation=[1]
                     /codon_start=1
                     /product="platelet glycoprotein V"
                     /protein_id="CAA93441.1"
                     /db_xref="PID:e222202"
                     /db_xref="PID:g2104845"
                     /db_xref="GI:2104845"
                     /db_xref="MGD:MGI:1096363"
                     /db_xref="SWISS-PROT:O08742"
                     /translation="MLRSALLSAVLPLLRAQPFPCPKTCKCVVRDAAQCSGGSVAHIA
                     ELGLPTNLTHILLFRMDQGILRNHSFSGMTVLQRQMLSDSHISAIDPGTFNDLVKLKT
                     LRLTRNKISRLPRAILDKMVLLEQLFLDHNALRDLDQNLFQQLRNLQELGLNQNQLSF
                     LPANLFSSLRELKLLDLSRNNLTHLPKGLLGAQVKLEKLLLYSNQLTSVDSGLLSNLG
                     ALTELRLERNHLRSVAPGAFDRLGNLSSLTLSGNLLESLPPALFLHVSSVSRLTLFEN
                     PLEELPDVLFGEMAGLRELWLNGTHLSTLPAAAFRNLSGLQTLGLTRNPRLSALPRGV
                     FQGLRELRVLGLHTNALAELRDDALRGLGHLRQVSLRHNRLRALPRTLFRNLSSLESV
                     QLEHNQLETLPGDVFAALPQLTQVLLGHNPWLCDCGLWRFLQWLRHHPDILGRDEPPQ
                     CRGPEPRASLSFWELLQGDPWCPDPRSLPLDPPTENALEAPVPSWLPNSWQSQTWAQL
                     VARGESPNNRLYWGLYILLLVAQAIIAAFIVFAMIKIGQLFRTLIREKLLLEAMGKSC
                     N"
     mat_peptide     1456. .3108
                     /standard_name="platelet GPV"
                     /function="von Willebrand factor receptor subunit,
                     thrombin substrate"
                     /product="platelet glycoprotein V"
     repeat_region   1558. .1629
                     /standard_name="LR domain"
                     /note="15 LR repeats of 24 amino acids except repeat 12 of
                     25 amino acids"
                     /citation=[1]
                     /rpt_family="leucine-rich"
                     /rpt_unit=1558. .1629
                     /function="structural element"
                     /rpt_type=TANDEM
     misc_feature    2836. .2838
                     /note="cleavage site"
                     /function="thrombin cleavage site"
     misc_feature    2962. .3036
                     /citation=[1]
                     /function="transmembrane domain"
     3'UTR           3112. .3584
                     /citation=[1]
     polyA_signal    3540. .3545
                     /citation=[1]
BASE COUNT      825 a    960 c    931 g    868 t
ORIGIN


alignment_scores:
            Quality:  163.00              Length:    358
              Ratio:  0.876                 Gaps:     20
 Percent Similarity:  51.955   Percent Identity: 24.022

alignment_block:
 US-09-068-804-14 x MMPLGPV   ..

 Align seg 1/1  to: MMPLGPV  from: 1  to: 3584

       5 LeuAsnAsnGlyAsnProValLeuAsnValGlyAlaSerGlyLeuThrTh 21
         |||::::::   ::::::   |||      :::   ::::::|||||
    2023 CTCACGTCTGTGGATTCGGGGCTGCTGAGCAACCTGGGCGCCCTGACTGA 2072

      21 rLeuPro.....................................A 24
          |||                                        |
    2073 GCTGCGGCTGGAGCGGAATCACCTCCGCTCCGTAGCCCCGGGTGCCTTCG 2122

      24 spArgLeuProProHisIleThrThrLeuValIleProAspAsnAsnLeu 40
         ||||||||      ::::::::::::|||::::::::::::|||   |||
```

```
2123 ACCGCCTC...GGAAACCTGAGCTCCTTGACTCTATCCGGAAACCTCCTG 2169

  41 ThrSerLeuProGluLeuProGluGlyLeuArgGluLeuGluValSerGl 57
2170 GAGTCTCTGCCG......CCCGCGCTCTTC......CTTCACGTGAGCAG 2207

  57 yAsnLeuGlnLeuThr.....SerLeuProSerLeuProGln. 69
2208 CGTGTCTCGGCTGACTCTGTTCGAGAACCCCCTGGAGGAGCTCCCGGACG 2257

  70 .....GlyLeuGlnLysLeuTrpAlaTyrAsnAsn 79
2258 TGTTGTTCGGGGAGATGGCCGGCCTGCGGGAGCTGTGGCTGAACGGCACC 2307

  80 TrpLeuAlaSerLeuProThrLeuPro.....ProGlyLeuGl 92
2308 CACCTGAGCACGCTGCCCGCCGCTGCCTTCCGCAACCTGAGCGGCTTGCA 2357

  92 yAspLeuAlaValSer...AsnAsnGlnLeuThrSerLeuPro...GluM 107
2358 GACGCTGGGGCTGACGCGGAACCCGCGCCTGAGCGCGCTCCCGCGCGGCG 2407

 107 etProProAlaLeuArgGluLeuArgValSerGlyAsnAsnLeuThrSer 123
2408 TGTTCCAGGGCCTACGGGAGCTGCGCGTGCTCGGCCTGCACACCAACGCC 2457

 124 LeuArgAlaLeuProSerGly.....LeuGlnLy 133
2458 CTGGCGGAGCTGCGGGACGACGCGCTGCGCGGCCTCGGGCACCTGCGCCA 2507

 133 sLeuTrpAlaTyrAsnAsnArgLeuThrSerLeuProGluMet..... 147
2508 GGTGTCGCTGCGCCACAACCGGCTGCGGGCCCTGCCCCGCACGCTCTTCC 2557

 148 .....SerProGlyLeuGlnGluLeuAspValSerHisAsnGlnLeuThr 162
2558 GCAACCTCAGCAGCCTCGAGAGCGTGCAGCTAGAGCACAACCAGCTGGAG 2607

 163 ArgLeuPro...GlnSerLeuThrGlyLeuSerSerAlaAlaArgValTy 178
2608 ACGCTGCCAGGAGACGTGTTCGCGGCTCTGCCCCAGCTGACCCAGGTCCT 2657

 178 rLeuAspGlyAsnProLeuSerValArgThrArgAspArgLeuCysGly. 194
2658 GCTGGGTCACAACCCC.....TGGCTCTGCGACT 2686

 195 .....HisHisTrpProPheArgHisGlnAspThrLeu 205
2687 GTGGCCTGTGGCGCTTCCTCCAGTGGCTGCGGCATCACCCGGACATCCTG 2736

 206 .....ArgTyrGlyGl 209
2737 GGCCGAGACGAGCCCCCGCAGTGCCGTGGCCCGGAGCCACGCGCCAGCCT 2786

 209 yAlaPheArg..... 212
2787 GTCGTTCTGGGAGCTGCTGCAGGGTGACCCGTGGTGCCCGGATCCTCGCA 2836

 213 .....ProArgGluAlaArgAlaLeuHisLeuAlaValAla 224
2837 GCCTGCCTCTCGACCCTCCAACCGAAAATGCTCTGGAAGCCCCGGTTCCG 2886

 225 AspTrpLeuThrSerAlaArgGluGlyGluAla..... 235
2887 TCCTGGCTGCCTAACAGCTGGCAGTCCCAGACGTGGGCCCAGCTGGTGGC 2936

 236 .....AlaGlnAlaAspArg.....TrpGlnAlaPheGlyLeu. 246
2937 CAGGGGTGAAAGTCCCAATAACAGGCTCTACTGGGGTCTTTATATTCTGC 2986

 247 .....GluAspAsnAlaAlaAlaPheSerLeuValLeuAspArg 259
2987 TTCTAGTAGCCCAGGCCATCATAGCCGCGTTC...ATCGTGTTTGCCATG 3033

 260 LeuArgGluThrGluAsnPheLys 267
3034 ATTAAAATCGGCCAGCTGTTTCGA 3057
```

```
seq_name: gb_ro:RNPLGPV

seq_documentation_block:
LOCUS       RNPLGPV      3573 bp    DNA             ROD       16-MAY-1997
DEFINITION  R.norvegicus platelet glycoprotein V gene.
ACCESSION   Z69594
NID         g2104855
VERSION     Z69594.1  GI:2104855
KEYWORDS    cis-acting element; platelet glycoprotein V; repeat region.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 3573)
  AUTHORS   Ravanat,C., Morales,M., Azorsa,D.O., Moog,S., Schuhler,S.,
            Grunert,P., Loew,D., Van Dorsselaer,A., Cazenave,J.P. and Lanza,F.
  TITLE     Gene cloning of rat and mouse platelet glycoprotein V:
            identification of megakaryocyte-specific promoters and
            demonstration of functional thrombin cleavage
  JOURNAL   Blood 89 (9), 3253-3262 (1997)
  MEDLINE   97275136
REFERENCE   2  (bases 1 to 3573)
  AUTHORS   Lanza,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-FEB-1996) Lanza F., Etablissement de Transfusion
            Sanguine de Strasbourg, INSERM U.311, 10, rue Spielmann,
            Strasbourg, FRANCE, 67065
FEATURES             Location/Qualifiers
     source          1. .3573
                     /organism="Rattus norvegicus"
                     /strain="Wistar"
                     /db_xref="taxon:10116"
                     /sex="Male"
                     /clone="lambda RV2"
                     /dev_stage="adult"
                     /tissue_type="liver"
                     /clone_lib="lambda DASH II genomic library"
     protein_bind    373. .378
                     /note="conserved in human and mouse genes"
                     /citation=[1]
                     /function="cis-acting element"
                     /bound_moiety="GATA-1"
     protein_bind    complement(410. .415)
                     /note="conserved in human and mouse genes"
                     /citation=[1]
                     /function="cis-acting element"
                     /bound_moiety="GATA-1"
     protein_bind    418. .424
                     /note="conserved in human and mouse genes"
                     /citation=[1]
                     /function="cis-acting element"
                     /bound_moiety="Ets-1"
     protein_bind    complement(439. .446)
                     /note="conserved in human and mouse genes"
                     /citation=[1]
                     /function="cis-acting element"
                     /bound_moiety="Ets-1"
     protein_bind    453. .463
                     /standard_name="Platelet GPV"
                     /citation=[1]
                     /function="cis-acting element"
                     /bound_moiety="NF-E2"
     prim_transcript 484. .>3573
                     /standard_name="platelet GPV"
                     /note="transcription start conserved in human and mouse
                     genes"
                     /citation=[1]
     exon            485. .513
```

```
                     /citation=[1]
                     /number=1
     intron          514. .1387
                     /citation=[1]
                     /number=1
     exon            1388. .3573
                     /citation=[1]
                     /number=2
     CDS             1390. .3093
                     /standard_name="platelet GPV"
                     /note="precursor"
                     /citation=[1]
                     /codon_start=1
                     /product="platelet glycoprotein V"
                     /protein_id="CAA93440.1"
                     /db_xref="PID:e222201"
                     /db_xref="PID:g2104856"
                     /db_xref="GI:2104856"
                     /db_xref="SWISS-PROT:O08770"
                     /translation="MLRSVLLSAVLSLVGAQPFPCPKTCKCVVRDAVQCSGGSVAHIA
                     ELGLPTNLTHILLFRMDRGVLQSHSFSGMTVLQRLMLSDSHISAIDPGTFNDLVKLKT
                     LRLTRNKISHLPRAILDKMVLLEQLFLDHNALRDLDQNLFQKLLNLRDLCLNQNQLSF
                     LPANLFSSLGKLKVLDLSRNNLTHLPQGLLGAQIKLEKLLLYSNRLMSLDSGLLANLG
                     ALTELRLERNHLRSIAPGAFDSLGNLSTLTLSGNLLESLPPALFLHVSWLTRLTLFEN
                     PLEELPEVLFGEMAGLRELWLNGTHLRTLPAAAFRNLSGLQTLGLTRNPLLSALPPGM
                     FHGLTELRVLAVHTNALEELPEDALRGLGRLRQVSLRHNRLRALPRTLFRNLSSLVTV
                     QLEHNQLKTLPGDVFAALPQLTRVLLGHNPWLCDCGLWPFLQWLRHHLELLGRDEPPQ
                     CNGPESRASLTFWELLQGDQWCPSSRGLPPDPPTENALKAPDPTQRPNSSQSWAWVQL
                     VARGESPDNRFYWNLYILLLIAQATIAGFIVFAMIKIGQLFRTLIREELLFEAMGKSS
                     N"
     sig_peptide     1390. .1437
     mat_peptide     1438. .3090
                     /function="von Willebrand factor receptor subunit,thrombin
                     substrate"
                     /product="platelet glycoprotein V"
     repeat_region   1540. .2622
                     /standard_name="LR domain"
                     /note="15 tandem LR repeats of 24 amino acids (except LR
                     repeat 12 of 25 amino acids)"
                     /citation=[1]
                     /rpt_family="leucine-rich"
                     /rpt_unit=1540. .1611
                     /function="structural element"
                     /rpt_type=TANDEM
     misc_feature    2817^2818
                     /note="site similar to human GPV and fibrinogen cleavage
                     sites"
                     /citation=[1]
                     /function="thrombin cleavage site"
     misc_feature    2944. .3021
                     /citation=[1]
                     /function="transmembrane domain"
     3'UTR           3094. .>3573
                     /note="lacks a polyadenylation signal"
                     /citation=[1]
BASE COUNT      798 a    962 c    937 g    876 t
ORIGIN


alignment_scores:
            Quality:  162.50              Length:    322
              Ratio:  1.055                 Gaps:     20
 Percent Similarity:  47.826   Percent Identity: 28.261

alignment_block:
 US-09-068-804-14 x RNPLGPV   ..

 Align seg 1/1  to: RNPLGPV  from: 1  to: 3573

      11 ValLeuAsnValGlyAlaSerGlyLeuThrThrLeuPro.......... 23
         ||||||:::::::::   :::::||||||   ||||||
    1909 GTGTTGGATTTATCGCGAAACAACCTGACCCACCTGCCCCAGGGATTACT 1958

      24 .....AspArgLeuP 27
                 :::||||||
    1959 TGGGGCGCAAATTAAGCTTGAGAAACTGCTGCTCTATTCAAACCGGCTCA 2008

      27 roProHisIleThrThrLeuValIleProAspAsnAsnLeuThrSerLeu 43
           :::      :::   |||:::         :::||||||   :::|||
    2009 TGTCTCTGGATTCGGGGCTGCTG.........GCCAACCTGGGCGCCCTG 2049

      44 ProGluLeuPro..................................... 47
         :::||||||
    2050 ACTGAGCTGCGGCTGGAGCGGAATCACCTTCGCTCCATCGCACCGGGTGC 2099

      48 .............GluGlyLeuArgGluLeuGluValSerGlyAsnLeuG 60
                        :::|||      |||   :::|||||||||||||
    2100 CTTCGACAGCCTCGGAAACCTGAGCACCTTGACTCTGTCCGGAAACCTC. 2148

      60 lnLeuThrSerLeuProSer.............................. 66
           |||   |||||||||:::
    2149 ..CTGGAGTCTCTGCCGCCCGCGCTCTTCCTTCATGTGAGCTGGCTGACT 2196

      67 ................................LeuProGln........ 69
                                         ||||||:::
    2197 CGGCTGACCCTGTTCGAGAACCCTCTGGAGGAGCTCCCGGAGGTGCTGTT 2246

      70 .............GlyLeuGlnLysLeuTrpAlaTyrAsnAsnTrpLeuA 82
                      |||||||:::  ||||||     ::::::   |||
    2247 CGGGGAGATGGCCGGCCTGCGGGAGCTGTGGCTGAACGGCACCCACCTGC 2296

      82 laSerLeuProThrLeuPro.............ProGlyLeuGlyAspLeu 94
           :::|||||||:::   :::            :::|||||||      |||
    2297 GCACGCTGCCCGCCGCCGCCTTCCGCAACCTGAGCGGCCTGCAGACGCTG 2346

      95 AlaValSerAsnAsnGlnLeuThrSerLeuProGluMetProPro..... 109
         :::::::::   |||:::|||      |||:::::::::|||||
    2347 GGGCTGACACGGAACCCGCTC......CTGAGCGCGCTCCCGCCGGGCAT 2390

     110 .......AlaLeuArgGluLeuArgValSerGlyAsnAsnLeuThrSerL 124
               :::|||   ||||||||||||   :::   :::   ::::::|
    2391 GTTCCACGGCCTGACGGAGCTGCGCGTGCTGGCCGTGCACACCAATGCCC 2440

     124 euArgAlaLeuPro.............SerGlyLeuGlnLysLeuTrpAla 136
          ||   :::||||||             ||||||   ::::|||:::
    2441 TGGAGGAGCTGCCGGAGGACGCGCTGCGCGGCCTCGGGCGCCTGCGCCAG 2490

     137 Tyr.........AsnAsnArgLeuThrSerLeuProGluMet........ 147
                   :::|||||||||   :::||||||
    2491 GTGTCGCTGCGCCACAACCGGTTGCGGGCCCTGCCCCGCACGCTCTTCCG 2540

     148 ....SerProGlyLeuGlnGluLeuAspValSerHisAsnGlnLeuThrA 163
               ::::::|||      :::::::::   ||||||||||||:::
    2541 TAACCTCAGCAGCCTCGTGACGGTGCAGCTTGAGCATAACCAGCTGAAGA 2590

     163 rgLeuPro...GlnSerLeuThrGlyLeuSerSerAlaAlaArgValTyr 178
           ||||||   :::   :::::::::|||:::      :::||||||
    2591 CCCTGCCAGGAGACGTGTTCGCGGCTCTGCCCCAACTGACCCGGGTCCTG 2640

     179 LeuAspGlyAsnProLeuSerValArgThrArgAspArgLeuCysGly.. 194
         |||:::   ||||||                    :::||||||:::
    2641 CTCGGTCACAACCCC.....................TGGCTCTGCGACTG 2669

     195 .HisHisTrpProPhe.............ArgHisGlnAspThrLeuArgT 207
              |||||||||              ||||||:::
    2670 TGGCCTGTGGCCCTTCCTCCAGTGGCTGCGGCATCACCT........... 2708

     207 yrGlyGlyAlaPheArgProArgGluAlaArgAlaLeuHisLeuAlaVal 223
           ||||||   ||||||||||   |||   |||::::::
    2709 .....GGAGCTCCTAGGCCGAGATGAGCCTCCGCAGTGCAA......... 2744

     224 AlaAspTrpLeuThrSerAlaArgGluGlyGlu.........AlaAla.G 237
           |||         :::|||:::   :::        ||||||  |
    2745 ......TGGCCCAGAGTCACGCGCCAGCCTGACGTTCTGGGAGCTGCTGC 2788

     237 lnAlaAspArgTrp 241
```

```
        ||:::|||:::|||
    2789 AGGGTGACCAGTGG 2802

seq_name: gb_pat:AR025202

seq_documentation_block:
LOCUS       AR025202      906 bp    DNA             PAT       04-DEC-1998
DEFINITION  Sequence 1 from patent US 5798216.
ACCESSION   AR025202
NID         g3977830
VERSION     AR025202.1  GI:3977830
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 906)
  AUTHORS   Ruggeri,Z.M. and Ware,J.L.
  TITLE     Antibodies and immunoassays for detecting platelet transfusion
            refractoriness
  JOURNAL   Patent: US 5798216-A 1 25-AUG-1998;
FEATURES             Location/Qualifiers
     source          1. .906
                     /organism="unknown"
BASE COUNT      212 a    282 c    220 g    192 t
ORIGIN


alignment_scores:
            Quality:  152.50            Length:    230
              Ratio:  1.292               Gaps:      8
 Percent Similarity:  51.304   Percent Identity: 28.261

alignment_block:
 US-09-068-804-14 x AR025202   ..

 Align seg 1/1  to: AR025202  from: 1  to: 906

      19 LeuThrThrLeuProAspArgLeuProProHisIleThrThrLeuValIl 35
         ||||||:::||||||      ||||||   ::::::|||:::|||   ::
      64 CTGACAGCGCTGCCTCCAGACCTGCCGAAAGACACAACCATCCTCCACCT 113

      35 eProAspAsnAsnLeu.............................T 41
         :::::::|||   |||                             |
     114 GAGTGAGAACCTCCTGTACACCTTCTCCCTGGCAACCCTGATGCCTTACA 163

      41 hrSerLeuProGluLeuProGluGly.........LeuArgGluLeuGlu 54
         ||   |||::::::|||      :::         |||      |||:::
     164 CTCGCCTCACTCAGCTGAACCTAGATAGGTGCGAGCTCACCAAGCTCCAG 213

      55 ValSerGlyAsnLeuGlnLeuThrSerLeuProSerLeuProGlnGlyLe 71
         |||:::|||:::|||
     214 GTCGATGGGACGCTG................................. 228

      71 uGlnLysLeuTrpAlaTyrAsnAsnTrpLeuAlaSerLeuProThrLeuP 88

     228 ..................................................... 228

      88 roProGlyLeuGlyAspLeuAlaValSerAsnAsnGlnLeuThrSerLeu 104
            |||   ||||||   |||:::::|||:::|||||||||   ||||||
     229 ..CCAGTGCTGGGGACCCTGGATCTATCCCACAATCAGCTGCAAAGCCTG 276

     105 ProGluMetProProAlaLeuArgGluLeu..................... 114
         |||   :::   ::::::|||   :::|||
     277 CCCTTGCTAGGGCAGACACTGCCTGCTCTCACCGTCCTGGACGTCTCCTT 326

     115 ....ArgValSerGlyAsnAsnLeuThrSerLeuArgAlaLeuProSerG 130
             |||::::::::::      |||   :::||||||:::|||   :::
     327 CAACCGGCTGACCTCGCTGCCTCTTGGTGCCCTGCGTGGTCTT...GGCG 373

     130 lyLeuGlnLysLeuTrpAlaTyrAsnAsnArgLeuThrSerLeuProGlu 146
           ||||||   |||         :::|||   |||::::::||||||
     374 AACTCCAAGAGCTCTACCTGAAAGGCAATGAGCTGAAGACCCTGCCCCCA 423

     147 ...........MetSerProGlyLeuGlnGluLeuAspValSerHisAs 159
                     :::|||   |||:::   |||:::::::::::||
     424 GGGCTCCTGATGCCCACACCCAAGCTGGAGAAGCTCAGTCTGGCTAACAA 473

     159 nGlnLeuThrArgLeuProGlnSer...LeuThrGlyLeuSerSerAlaA 175
         |:::||||||   ||||||   :::   |||:::||||||   :::   :
     474 CAACTTGACTGAGCTCCCCGCTGGGCTCCTGAATGGGCTGGAGAATCTCG 523

     175 laArgValTyrLeuAspGlyAsnProLeuSerValArgThrArgAspArg 191
         ::   :::   |||:::  |||:::|||  :::  :::::::::
     524 ACACCCTTCTCCTCCAAGAGAACTCGCTGTATACAATACCAAAGGGCTTT 573

     192 LeuCysGlyHisHisTrpProPheArgHisGlnAspThrLeuArgTyrGl 208
         :::   :::|||      ||||||
     574 TTTGGGTCCCACCTCCTGCCTTTTGC..................... 599

     208 yGlyAlaPheArgProArgGluAlaArgAlaLeuHisLeu 221
               |||   |||||||||:::   :::::::::|||
     600 .......TTTTCTCCACGGGAACCCCTGGTTATGCAACTG 632

seq_name: gb_pat:I12851

seq_documentation_block:
LOCUS       I12851        906 bp    DNA             PAT       07-JUL-1995
DEFINITION  Sequence 1 from patent US 5427939.
ACCESSION   I12851
NID         g910233
VERSION     I12851.1  GI:910233
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 906)
  AUTHORS   Ruggeri,Z.M. and Ware,J.L.
  TITLE     Expression vectors encoding alloantigens of glycoprotein Ib.alpha
  JOURNAL   Patent: US 5427939-A 1 27-JUN-1995;
FEATURES             Location/Qualifiers
     source          1. .906
                     /organism="unknown"
BASE COUNT      212 a    282 c    220 g    192 t
ORIGIN


alignment_scores:
            Quality:  152.50            Length:    230
              Ratio:  1.292               Gaps:      8
 Percent Similarity:  51.304   Percent Identity: 28.261

alignment_block:
 US-09-068-804-14 x I12851   ..

 Align seg 1/1  to: I12851  from: 1  to: 906

      19 LeuThrThrLeuProAspArgLeuProProHisIleThrThrLeuValIl 35
         ||||||:::||||||      ||||||   ::::::|||:::|||   ::
      64 CTGACAGCGCTGCCTCCAGACCTGCCGAAAGACACAACCATCCTCCACCT 113

      35 eProAspAsnAsnLeu.............................T 41
         :::::::|||   |||                             |
     114 GAGTGAGAACCTCCTGTACACCTTCTCCCTGGCAACCCTGATGCCTTACA 163

      41 hrSerLeuProGluLeuProGluGly.........LeuArgGluLeuGlu 54
         ||   |||::::::|||      :::         |||      |||:::
     164 CTCGCCTCACTCAGCTGAACCTAGATAGGTGCGAGCTCACCAAGCTCCAG 213

      55 ValSerGlyAsnLeuGlnLeuThrSerLeuProSerLeuProGlnGlyLe 71
         |||:::|||:::|||
     214 GTCGATGGGACGCTG................................. 228

      71 uGlnLysLeuTrpAlaTyrAsnAsnTrpLeuAlaSerLeuProThrLeuP 88

     228 ..................................................... 228
```

```
 88 roProGlyLeuGlyAspLeuAlaValSerAsnAsnGlnLeuThrSerLeu 104
    |||   ||||||   |||:::::::|||:::|||||||||   ||||||
229 ..CCAGTGCTGGGGACCCTGGATCTATCCCACAATCAGCTGCAAAGCCTG 276

105 ProGluMetProProAlaLeuArgGluLeu..................... 114
    |||   :::   ::::::|||   :::|||
277 CCCTTGCTAGGGCAGACACTGCCTGCTCTCACCGTCCTGGACGTCTCCTT 326

115 ....ArgValSerGlyAsnAsnLeuThrSerLeuArgAlaLeuProSerG 130
        |||::::::::::      |||   :::||||||:::|||   :::
327 CAACCGGCTGACCTCGCTGCCTCTTGGTGCCCTGCGTGGTCTT...GGCG 373

130 lyLeuGlnLysLeuTrpAlaTyrAsnAsnArgLeuThrSerLeuProGlu 146
    ||||||   |||         :::|||   |||:::::|||||
374 AACTCCAAGAGCTCTACCTGAAAGGCAATGAGCTGAAGACCCTGCCCCCA 423

147 ............MetSerProGlyLeuGlnGluLeuAspValSerHisAs 159
                :::|||   |||:::   |||:::::::::::||
424 GGGCTCCTGATGCCCACACCCAAGCTGGAGAAGCTCAGTCTGGCTAACAA 473

159 nGlnLeuThrArgLeuProGlnSer...LeuThrGlyLeuSerSerAlaA 175
    |:::||||||   ||||||   :::   |||:::||||||   :::   :
474 CAACTTGACTGAGCTCCCCGCTGGGCTCCTGAATGGGCTGGAGAATCTCG 523

175 laArgValTyrLeuAspGlyAsnProLeuSerValArgThrArgAspArg 191
    ::   :::   |||:::   |||:::|||   :::   :::::::::
524 ACACCCTTCTCCTCCAAGAGAACTCGCTGTATACAATACCAAAGGGCTTT 573

192 LeuCysGlyHisHisTrpProPheArgHisGlnAspThrLeuArgTyrGl 208
    :::   :::|||      ||||||
574 TTTGGGTCCCACCTCCTGCCTTTTGC..................... 599

208 yGlyAlaPheArgProArgGluAlaArgAlaLeuHisLeu 221
        |||   |||||||||:::   :::::::::|||
600 .......TTTTCTCCACGGGAACCCCTGGTTATGCAACTG 632
```

This Page Blank (uspto)

```
OM of: US-09-068-804-14 to: GenEmbl:*   out_format : pfs

Date: Oct 5, 1999  8:02 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=x1p
 -Q=/cgn2_1/USPTO_spool/US09068804/runat_05101999_145134_10315/app_query.fasta.1
 -DB=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=pam150 -TRANS=human40.cdi
 -LIST=45 -DOCALIGN=200 -THR_SCORE=escore -ALIGN=15 -MODE=LOCAL
 -OUTFMT=pfs -NORM=stat -USER=US09068804 -NCPU=6 -ICPU=3 -WAIT
 -THREADS=1

Search information block:
Query: US-09-068-804-14
Query length: 513
Database: GenEmbl:*
Database sequences: 679419
Database length: 1590154680
Search time (sec): 1424.920000


score_list:
Sequence              Strd Orig    ZScore    EScore Len  ! Documentation  ..
gb_ba1:SHFIPAHZ         +  875.50 1170.05  3.1e-57   1992  ! M76445 Shigella flexneri ipaH
gb_ba1:SHFIPAH          +  853.00 1136.40  2.3e-55   2900  ! M32063 S.flexner invasion plas
gb_ba2:AF047365         +  821.50 1098.16  3.2e-53   1822  ! AF047365 Shigella flexneri pla
gb_ba2:AE000074         +  619.00  807.14  5.1e-37  14263  ! AE000074 Rhizobium sp. NGR234
gb_ba1:SHFIPAHX         +  602.50  806.28  5.7e-37   1312  ! M76443 Shigella flexneri ipaH
gb_ba1:SHFIPAHY         +  602.50  805.86  6.0e-37   1375  ! M76444 Shigella flexneri ipaH
gb_ba1:YEPYOPM          +  272.00  360.81  3.7e-12   1397  ! M25810 Y.pestis outer membrane
gb_ba2:AF102990         +  286.00  344.66  2.9e-11  69673  ! AF102990 Yersinia enterocolit
gb_ba2:AF053946         +  266.00  317.63  9.4e-10  70504  ! AF053946 Yersinia pestis plas
gb_ba2:AF074612         +  266.00  317.63  9.4e-10  70559  ! AF074612 Yersinia pestis plas
gb_pr1:AB016816         +  170.50  216.87   0.0004   3159  ! AB016816 Homo sapiens MASL1 mR
gb_ro:MMPLGPV           +  163.00  205.64   0.0016   3584  ! Z69595 M.musculus platelet gly
gb_ro:RNPLGPV           +  162.50  205.00   0.0018   3573  ! Z69594 R.norvegicus platelet g
gb_pat:AR025202         +  152.50  203.81   0.0021    906  ! AR025202 Sequence 1 from patent
gb_pat:I12851           +  152.50  203.81   0.0021    906  ! I12851 Sequence 1 from patent U
gb_pl1:HALRPR           +  156.00  201.83   0.0027   1915  ! X81997 H.annuus SF17 mRNA for
gb_pr1:D63481           +  158.00  196.05   0.0056   4933  ! D63481 Human mRNA for KIAA0147
gb_ro:MMRSP1P33         -  149.00  195.69   0.0059   1327  ! X63039 M.musculus RSP-1 mRNA f
gb_pr1:HUMGPIBA         +  152.50  194.80   0.0066   2480  ! J02940 Human platelet glycoprc
gb_pr2:HSOMGPAA         +  148.50  192.25   0.0091   1808  ! X51694 H.sapiens OMGP mRNA for
gb_om:CFU19489          +  149.50  190.58   0.0113   2531  ! U19489 Canis familiaris glycop
gb_pr1:HUMRSU1A         +  148.00  189.84   0.0124   2194  ! L12535 Human RSU-1/RSP-1 mRNA,
gb_pr2:S83462           +  145.50  188.16   0.0154   1818  ! S83462 ALS=85 kda insulin-like
gb_pr1:HUMOMGPA         +  148.50  187.79   0.0161   2973  ! M63623 Human oligodendrocyte-m
gb_pr1:HUMGPIBAA        +  152.50  186.80   0.0183   6062  ! M22403 Human blood platelet me
gb_pr1:HSOMGP           +  148.50  185.87   0.0206   3686  ! X57436 H.sapiens gene for olig
gb_pat:AR012319         +  139.50  185.20   0.0225   1026  ! AR012319 Sequence 1 from paten
gb_pat:I27828           +  139.50  185.20   0.0225   1026  ! I27828 Sequence 1 from patent
gb_om:OCU03394          +  137.50  183.07   0.0295    964  ! U03394 Oryctolagus cuniculus de
gb_pr2:HUMCSPG1A        +  139.50  180.28   0.0422   1778  ! M14219 Human chondroitin/derma
gb_ro:S67043            +  144.50  179.14   0.0489   4286  ! S67043 oligodendrocyte-myelin
gb_om:S76584            +  137.50  178.80   0.0510   1553  ! S76584 decorin=proteoglycan [r
gb_pr1:HUMIGFACID       +  139.50  178.68   0.0518   2125  ! M86826 Human IGF binding prote
gb_in2:AF068919         +  137.50  178.09   0.0559   1680  ! AF068919 Caenorhabditis elegan
gb_pr3:AF068920         +  136.50  176.39   0.0695   1749  ! AF068920 Homo sapiens Ras-bind
gb_in1:AF054827         +  137.50  175.91   0.0740   2145  ! AF054827 Caenorhabditis elegan
gb_om:RABCD14A          +  133.00  175.67   0.0762   1119  ! M85233 Rabbit CD14 mRNA, compl
gb_pr3:U01184           +  141.50  175.48   0.0781   4105  ! U01184 Human homolog of D. mel
gb_om:D16545            +  133.00  174.65   0.0869   1255  ! D16545 Oryctolagus cuniculus C
gb_pr3:AF054828         +  136.50  173.88   0.0960   2316  ! AF054828 Homo sapiens leucine-
gb_ro:AF061443          +  139.00  173.40   0.1021   3559  ! AF061443 Rattus norvegicus G p
gb_ro:AF068921          +  133.50  172.35   0.1167   1749  ! AF068921 Mus musculus Ras-bind
gb_om:AF038127          +  131.50  171.99   0.1223   1348  ! AF038127 Equus caballus der
gb_om:AF140270          +  131.00  171.12   0.1367   1378  ! AF140270 Sus scrofa decorin
gb_ro:RNU66707          +  139.50  170.83   0.1418   5109  ! U66707 Rattus norvegicus de

seq_name: gb_ba1:SHFIPAHZ

seq_documentation_block:
LOCUS       SHFIPAHZ     1992 bp    DNA              BCT       23-OCT-1991
DEFINITION  Shigella flexneri ipaH gene, 5'end.
ACCESSION   M76445
NID         g152751
VERSION     M76445.1  GI:152751
KEYWORDS    LRG-like protein; ipaH gene.
SOURCE      Shigella flexneri (strain 5) DNA.
  ORGANISM  Shigella flexneri
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Shigella.
REFERENCE   1  (bases 1 to 1992)
  AUTHORS   Hartman,A.B., Venkatesan,M., Oaks,E.V. and Buysse,J.M.
  TITLE     Sequence and molecular characterization of a multicopy invasion
            plasmid antigen gene, ipaH, of Shigella flexneri
  JOURNAL   J. Bacteriol. 172, 1905-1915 (1990)
  MEDLINE   90202708
REFERENCE   2  (bases 1 to 1992)
  AUTHORS   Venkatesan,M.M., Buysse,J.M. and Hartman,A.B.
  TITLE     Sequence variation in two ipaH genes of Shigella flexneri 5 and
            homology to the LRG-like family of proteins
  JOURNAL   Mol. Microbiol. 5, 2435-2446 (1991)
  MEDLINE   92167809
FEATURES             Location/Qualifiers
     source          1. .1992
                     /organism="Shigella flexneri"
                     /strain="5"
                     /db_xref="taxon:623"
     gene            1. .1725
                     /gene="ipaH"
     CDS             1. .1725
                     /gene="ipaH"
                     /codon_start=1
                     /transl_table=11
                     /product="antigen"
                     /protein_id="AAA26528.1"
                     /db_xref="PID:g152752"
                     /db_xref="GI:152752"
                     /translation="MKPINNHSFFRSLCGLSCISRLSVEEQCTRDYHRIWDDWAREGI
                     TTENRIQAVRLLKICLDTREPVLNLSLLKLRSLPPLPLHIRELNISNNELISLPENSP
                     LLTELHVNGNNLNILPTLPSQLIKLNISFNRNLSCLPSLPPYLQSLSARFNSLETLPE
                     LPSTLTILRIEGNRLTVLPELPHRLQELFVSGNRLQELPEFPQSLKYLKVGENQLRRL
                     SRLPQELLALDVSNNLLTSLPENIITLPICTNVNISGNPLSTRVLQSLQRLTSSPDYH
                     GPQIYFSMSDGQQNTLHRPLADAVTAWFPENKQSDVSQIWHAFEHEEHANTFSAFLDR
                     LSDTVSARNTSGFREQVAAWLEKLSASAELRQQSFAVAADATESCEDRVALTWNNLRK
                     TLLVHQASEGLFDNDTGALLSLGREMFRLEILEDIARDKVRTLHFVDEIEVYLAFQTM
                     LAEKLQLSTAVKEMRFYGVSGVTANDLRTAEAMVRSREENEFTDWFSLWGPWHAVLKR
                     TEADRWAQAEEQKYEMLENEYSQRVADRLKASGLSGDADAQREAGAQVMRETEQQIYR
                     QLTDEVLA"
BASE COUNT      529 a    500 c    464 g    499 t
ORIGIN


alignment_scores:
            Quality:  875.50              Length:    547
              Ratio:  2.353                 Gaps:     10
 Percent Similarity:  68.007    Percent Identity: 39.305

alignment_block:
 US-09-068-804-14 x SHFIPAHZ   ..

 Align seg 1/1  to: SHFIPAHZ  from: 1  to: 1992

       1 MetArgAspCysLeuAsnAsnGlyAsnProValLeuAsnValGlyAlaSe 17
         ::::::   ||||||::::::   :::||||||||||||||::::::
     166 TTGAAAATATGTCTGGATACCCGGGAGCCTGTTCTCAATTTAAGCTTACT 215
```

```
 17 rGlyLeuThrThrLeuProAspArgLeuProProHisIleThrThrLeuV 34
216 GAAACTACGTTCTTTACCA...CCACTCCCTTTGCATATACGTGAACTTA 262

 34 alIleProAspAsnAsnLeuThrSerLeuProGluLeuProGluGlyLeu 50
263 ATATTTCCAACAATGAGTTAATCTCCCTACCTGAAAATTCTCCGCTTTTG 312

 51 ArgGluLeuGluValSerGlyAsnLeuGlnLeuThrSerLeuProSerLe 67
313 ACAGAACTTCATGTAAATGGTAAC...AACTTGAATATACTCCCGACACT 359

 67 uProGlnGlyLeuGlnLysLeuTrp...AlaTyrAsnAsnTrpLeuAlaS 83
360 TCCATCTCAACTGATTAAGCTTAATATTTCATTCAATCGAAATTTGTCAT 409

 83 erLeuProThrLeuProProGlyLeuGlyAspLeuAlaValSerAsnAsn 99
410 GTCTGCCATCATTACCACCATATTTACAATCACTCTCGGCACGTTTTAAT 459

100 GlnLeuThrSerLeuProGluMetProProAlaLeuArgGluLeuArgVa 116
460 AGTCTGGAGACGTTACCAGAGCTTCCATCAACGCTAACAATATTACGTAT 509

116 lSerGlyAsnAsnLeuThrSerLeuArgAlaLeuProSerGlyLeuGlnL 133
510 TGAAGGTAATCGCCTTACTGTCTTGCCTGAATTGCCTCATAGACTACAAG 559

133 ysLeuTrpAlaTyrAsnAsnArgLeuThrSerLeuProGlu..... 146
560 AACTCTTTGTTTCCGGCAACAGACTACAGGAACTACCAGAATTTCCTCAG 609

146 ..... 146
610 AGCTTAAAATATTTGAAGGTAGGTGAAAATCAACTACGCAGATTATCCAG 659

147 .MetSerProGlyLeuGlnGluLeuAspValSerHisAsnGlnLeuThrA 163
660 ATTACCGCAAGAACTATTGGCTCTGGATGTTTCCAATAACCTACTAACTT 709

163 rgLeuProGlnSerLeuThrGlyLeuSerSerAlaAlaArgValTyrLeu 179
710 CATTACCCGAAAATATAATCACATTGCCCATTTGTACGAATGTTAACATT 759

180 AspGlyAsnProLeuSerValArgThrArgAspArgLeu..... 192
760 TCAGGGAATCCATTGTCGACTCGCGTTCTGCAATCCCTGCAAAGATTAAC 809

193 .....CysGlyHisHisTrpProPheArg..... 200
810 CTCTTCGCCGGACTACCACGGCCCGCAGATTTACTTCTCCATGAGTGACG 859

201 ..HisGlnAspThrLeuArgTyrGlyGlyAlaPheArgProArgGluAla 216
860 GACAACAGAATACACTCCAT..... 879

217 ArgAlaLeuHisLeuAlaValAlaAspTrpLeuThrSerAlaArgGluGl 233
880 CGCCCCCTGGCTGATGCCGTGACAGCATGGTTCCCGGAAAACAAACAATC 929

233 yGluAlaAlaGlnAlaAspArgTrpGlnAlaPheGlyLeuGluAspAsnA 250
930 TGATGTATCACAGATA......TGGCATGCTTTTGAACATGAAGAGCATG 973

250 laAlaAlaPheSerLeuValLeuAspArgLeuArgGluThrGluAsnPhe 266
974 CCAACACCTTTTCCGCGTTCCTTGACCGCCTTTCCGATACCGTCTCTGCA 1023

267 LysLysAspAlaGlyPheLysAlaGlnIleSerSerTrpLeuThrGlnLe 283
1024 CGCAATACCTCCGGATTCCGTGAACAGGTCGCTGCATGGCTGGAAAAACT 1073

283 uAlaGluAspAlaAlaLeuArgAlaLysThrPheAlaMetAlaThrGluA 300

1074 CAGTGCCTCTGCGGAGCTTCGACAGCAGTCTTTCGCTGTTGCTGCTGATG 1123

300 laThrSerThrCysGluAspArgValThrHisAlaLeuHisGlnMetAsn 316
1124 CCACTGAGAGCTGTGAGGACCGTGTCGCGCTCACATGGAACAATCTCCGG 1173

317 AsnValGlnLeuValHisAsnAlaGluLysGlyGluTyrAspAsnAsnLe 333
1174 AAAACCCTCCTGGTCCATCAGGCATCAGAAGGCCTTTTCGATAATGATAC 1223

333 uGlnGlyLeuValSerThrGlyArgGluMetPheArgLeuAlaThrLeuG 350
1224 CGGCGCTCTGCTCTCCCTGGGCAGGGAAATGTTCCGCCTCGAAATTCTGG 1273

350 luGlnIleAlaArgGluLysAlaGlyThrLeuAlaLeuValAspAspVal 366
1274 AGGACATTGCCCGGGATAAAGTCAGAACTCTCCATTTTGTGGATGAGATA 1323

367 GluValTyrLeuAlaPheGlnAsnLysLeuLysGluSerLeuGluLeuTh 383
1324 GAAGTCTACCTGGCCTTCCAGACCATGCTCGCAGAGAAACTTCAGCTCTC 1373

383 rSerValThrSerGluMetArgPhePheAspValSerGlyValThrValS 400
1374 CACTGCCGTGAAGGAAATGCGTTTCTATGGCGTGTCGGGAGTGACAGCAA 1423

400 erAspLeuGlnAlaAlaAspValGlnValLysThrAlaGluAsnSerGly 416
1424 ATGACCTCCGCACTGCCGAAGCTATGGTCAGAAGCCGTGAAGAGAATGAA 1473

417 PheSerLysTrpIleLeuGlnTrpGlyProLeuHisSerValLeuGluAr 433
1474 TTTACGGACTGGTTCTCCCTCTGGGGACCATGGCATGCTGTACTGAAGCG 1523

433 gLysValProGluArgPheAsnAlaLeuArgGluLysGlnIleSerAspT 450
1524 TACGGAAGCTGACCGCTGGGCGCAGGCAGAAGAGCAGAAGTATGAGATGC 1573

450 yrGluAspThrTyrArgLysLeuTyrAspGluValLeuLysSerSerGly 466
1574 TGGAGAATGAGTACTCTCAGAGGGTGGCTGACCGGCTGAAAGCATCAGGT 1623

467 LeuValAspAspThrAspAlaGluArgThrIleGlyValSerAlaMetAs 483
1624 CTGAGCGGTGATGCGGATGCGCAGAGGGAAGCCGGTGCACAGGTGATGCG 1673

483 pSerAlaLysLysGluPheLeuAspGlyLeuArgAlaLeuValAspGluV 500
1674 TGAGACTGAACAGCAGATT.....TACCGTCAGCTGACTGACGAGG 1714

500 alLeuGly.SerTyrLeuThrAlaArgTrpArgLeuAsn 512
1715 TACTGGCCTGAACCGCCCCGGGAATCCTGGAGACTAAAC 1753
```

seq_name: gb_bal:SHFIPAH

seq_documentation_block:

```
LOCUS       SHFIPAH      2900 bp    DNA             BCT       23-JAN-1991
DEFINITION  S.flexner invasion plasmid antigen H (ipaH) gene, complete cds.
ACCESSION   M32063
NID         g152746
VERSION     M32063.1  GI:152746
KEYWORDS    invasion plasmid antigen.
SOURCE      S.flexner (strain M90T-W; serotype 5) DNA, clone pWR390.
  ORGANISM  Shigella flexneri
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Shigella.
REFERENCE   1  (bases 1 to 2900)
  AUTHORS   Hartman,A.B., Venkatesan,M., Oaks,E.V. and Buysse,J.M.
  TITLE     Sequence and molecular characterization of a multicopy invasion
            plasmid antigen gene, ipaH, of Shigella flexneri
  JOURNAL   J. Bacteriol. 172, 1905-1915 (1990)
```

```
OM of: US-09-068-804-7 to: EST:*   out_format : pfs

Date: Oct 5, 1999  7:34 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn2_1/USPTO_spool/US09068804/runat_05101999_145134_10306/app_query.fasta.1
 -DB=EST -QFMT=fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=pam150 -TRANS=human40.cdi
 -LIST=45 -DOCALIGN=200 -THR_SCORE=escore -ALIGN=15 -MODE=LOCAL
 -OUTFMT=pfs -NORM=stat -USER=US09068804 -NCPU=6 -ICPU=3 -WAIT
 -THREADS=1

Search information block:
Query: US-09-068-804-7
Query length: 344
Database: EST:*
Database sequences: 2546578
Database length: 986266752
Search time (sec): 940.790000


score_list:
Sequence              Strd Orig    ZScore    EScore Len  ! Documentation  ..
gb_est29:AI542214      +    85.00   182.74   0.1155    619  ! AI542214 SD08362.5prime SD Dros
gb_est29:AI556365      -    80.50   173.66   0.3702    523  ! AI556365 UI-R-C2p-rg-f-01-0-UI.
gb_est29:AI552661      -    76.50   166.19   0.9643    427  ! AI552661 ml98g09.y1 Stratagene
gb_est6:W01887         +    76.00   163.85   1.30      471  ! W01887 za35h04.r1 Soares fetal
gb_est26:AI324162      +    77.50   162.43   1.56      735  ! AI324162 mh66a06.x1 Soares mous
gb_est21:AA937937      +    73.50   160.97   1.88      355  ! AA937937 oc08b01.s1 NCI_CGAP_GC
gb_est27:AI398188      +    75.00   160.58   1.98      506  ! AI398188 NCSM2F10T3 Subtracted
gb_est16:AA550027      -    75.50   159.99   2.14      593  ! AA550027 1106m3 gmbPfHB3.1, G.
gb_est27:AI456803      +    73.50   156.85   3.20      509  ! AI456803 LD36936.5prime LD Dros
gb_est28:AI488463      +    73.00   156.01   3.56      492  ! AI488463 EST246802 tomato ovary
gb_est19:AA777598      +    72.00   155.82   3.65      404  ! AA777598 zj03a10.s1 Soares_feta
gb_est28:AI483429      +    73.00   155.80   3.65      501  ! AI483429 EST249250 tomato ovary
gb_est28:AI532519      +    74.00   154.94   4.08      669  ! AI532519 SD04057.5prime SD Dros
gb_est1:T01512         +    70.50   154.43   4.36      331  ! T01512 wEST02233 Early embryo,
em_est22:HSM007904     +    74.50   154.18   4.50      796  ! AI043054 Homo sapiens mRNA; EST
gb_est17:AA648521      -    73.50   154.05   4.58      650  ! AA648521 ns22g08.s1 NCI_CGAP_GC
gb_est17:C72813        +    71.50   153.86   4.69      431  ! C72813 C72813 Rice panicle at f
gb_est27:AI405872      +    74.00   153.62   4.84      751  ! AI405872 GH26024.5prime GH Dros
gb_est23:AI116302      -    72.50   153.41   4.97      555  ! AI116302 uc18h09.r1 Soares mous
gb_est28:AI518165      +    71.00   152.58   5.53      433  ! AI518165 LD37602.5prime LD Dros
gb_est18:AA683841      -    70.50   152.04   5.92      408  ! AA683841 vr05e02.r1 Knowles Sol
gb_est19:AA812903      -    71.00   152.01   5.94      455  ! AA812903 ai81g01.s1 Soares_test
gb_est4:H51273         -    71.00   151.99   5.96      456  ! H51273 yp83d06.s1 Soares fetal
gb_est24:AI173210      +    72.00   151.68   6.20      580  ! AI173210 uh17f01.r1 Barstead mc
gb_est27:AI405625      +    72.00   151.59   6.28      585  ! AI405625 GH25711.5prime GH Dros
gb_est28:AI516006      +    71.00   151.21   6.59      488  ! AI516006 LD41737.5prime LD Dros
gb_est30:AI654251      -    70.00   151.18   6.61      395  ! AI654251 tq89c05.x1 NCI_CGAP_Ov
gb_est25:AI296200      +    72.00   151.18   6.61      606  ! AI296200 LP10036.5prime LP Dros
gb_est6:W10801         -    69.00   150.79   6.95      330  ! W10801 ma47h11.r1 Soares mouse
gb_est23:AI108018      +    72.00   150.56   7.16      640  ! AI108018 GH05942.5prime GH Dros
gb_est29:AI568358      +    71.00   150.51   7.21      519  ! AI568358 th13c07.x1 NCI_CGAP_CL
gb_est26:AI354026      +    70.50   150.42   7.29      470  ! AI354026 zeh1296.seq.F Zebrafis
gb_est17:AA617917      -    71.00   150.20   7.49      533  ! AA617917 nq05h08.s1 NCI_CGAP_Lu
gb_est25:AU039891      +    71.50   150.04   7.66      602  ! AU039891 AU039891 Dictyostelium
gb_est17:AA640063      +    69.50   149.79   7.90      401  ! AA640063 np40c11.s1 NCI_CGAP_Lu
gb_est17:AA637189      +    70.50   149.18   8.55      524  ! AA637189 vr31c10.r1 Barstead mc
gb_est27:AI461045      -    68.50   149.10   8.64      344  ! AI461045 sa73c04.y1 Gm-c1004 Gl
gb_est8:AA041189       -    70.00   149.00   8.74      478  ! AA041189 zf08a08.r1 Soares_feta
gb_est27:AI444066      +    69.50   148.73   9.06      440  ! AI444066 sa31h03.y1 Gm-c1004 Gl
gb_est10:AA153489      +    72.00   148.56   9.25      762  ! AA153489 mr89a08.r1 Stratagene
gb_est31:F25270        +    67.50   148.49   9.34      293  ! F25270 HSPD12192 HM3 Homo sapie
gb_est5:N35204         +    68.50   147.84  10.15      384  ! N35204 yx84e12.r1 Soares meland
gb_est29:AI590940      -    70.00   147.50  10.60      545  ! AI590940 tw89c07.x1 NCI_CGAP
gb_est19:AA441170      -    70.50   147.18  11.05      624  ! AA441170 LD16071.5prime LD D
gb_est19:AA752186      +    70.50   146.87  11.49      641  ! AA752186 96BS0260 Rice Immat

seq_name: gb_est29:AI542214

seq_documentation_block:
LOCUS       AI542214      619 bp    mRNA            EST       22-MAR-1999
DEFINITION  SD08362.5prime SD Drosophila melanogaster Schneider L2 cell culture
            pOT2 Drosophila melanogaster cDNA clone SD08362 5prime similar to
            L07943: pum FBgn0003165 PID:g158191 SWISS-PROT:P25822, mRNA
            sequence.
ACCESSION   AI542214
NID         g4459587
VERSION     AI542214.1  GI:4459587
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 619)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     On Mar 10, 1998 this sequence version replaced gi:2948834.

            Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 83  row: F  column: 2
            High quality sequence stop: 411.
FEATURES             Location/Qualifiers
     source          1. .619
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="SD08362"
                     /clone_lib="SD Drosophila melanogaster Schneider L2 cell
                     culture pOT2"
                     /lab_host="DH5-alpha"
                     /note="Vector: pOT2; Site_1: EcoRI; Site_2: XhoI; Sized
                     fractionated cDNAs were directly ligated into pOT2.
                     Plasmid cDNA library."
BASE COUNT      106 a    231 c    194 g     88 t
ORIGIN


alignment_scores:
           Quality:   85.00              Length:    163
             Ratio:  0.895                Gaps:      6
 Percent Similarity:  58.282   Percent Identity: 24.540

alignment_block:
 US-09-068-804-7 x AI542214   ..

 Align seg 1/1  to: AI542214  from: 1  to: 619

      11 ProHisProGlyIleValAlaGluArgProGlnThrProSer......Al 25
         |||:::::|||:::   :::   ||||||   |||||||||       ::
      63 CCGCAACAGGGAACGCAGCCGCGCCGCCCCCTCACCCCCTCGCAGCAGGG 112

      25 aSerGluHisValGluThrAlaValValProSerThrThrGluHisArgG 42
         ::::|||:::          |||:::|||:::      :::|||   |
     113 TGCCGAGAATCAGCCGTATCAGGTCATCCCGGCATTCCTCGATCACACGG 162

      42 lyThrAspIleIleSerLeuSerGlnAlaAlaThrLysIleHis..... 56
         ||:::   :::::::::   :::::::::::::|||   ::::::
     163 GCTCCTTGCTGATGGGAGGACCTCGCACCGGGACGCCGATGCGTCTGGTT 212
```

```
     56 ..................................................  56

    213 AGCCCCGCCCCCGTTCTGGTGCCCCCGGGCGCTACCCGTGCCGGCCCCCC 262

     57 .......GlnAlaGlnGlnThrLeuGlnSerThrProProIleSerGluG  71
                |||::::::|||      |||:::   |||:::::::::::::
    263 GCCCCCGCAGGGCCCACAGCTGTATCAGCCGCAGCCGCAGACGGCCCAAC 312

     71 luAsnAsnAspGluArgThrLeuAlaArgGlnGlnLeuThrSerSerLeu  87
        ::|||               |||      ||||||      ||||||:::
    313 AGAAT..............CTCTACTCGCAGCAGAATGGATCCAGTGTC 347

     88 AsnAlaLeuAlaLysSerGlyValSerLeuSerAlaGluGlnAsnGluAs 104
        ::::::||||||   :::      ||||||::::::
    348 GGAGGCCTCGCCTTGAACACGAGCTCGTTGACGGGTCGC.......... 386

    104 nLeuArgSerAlaPheSerAlaProThrSerAlaLeuPheSerAlaSerP 121
           |||:::::::|||:::   :::|||||||||   ||||||:::|||:
    387 ....CGCGACTCCTTCGACCGCAGCACCTCCGCC...TTCAGTCCCTCGA 429

    121 roMetAlaGlnProArgThrThrIleSerAspAla......GluIleTrp 135
        ::|||:::   :::   :::   ::::::::::|||      ::::::
    430 CCATGGACTACACCAGCAGCGGTGTGGCAGCGGCCGCCAATGCGGTGAAC 479

    136 AspMetValSerGlnAsnIleSerAlaIleGlyAspSer 148
        :::   |||:::|||:::   :::|||   :::::::::
    480 AGCACAGTGGCCCAGGCAGCAGCAGCTGCCGCAGCAGCC 518

seq_name: gb_est29:AI556365

seq_documentation_block:
LOCUS       AI556365      523 bp    mRNA            EST       23-MAR-1999
DEFINITION  UI-R-C2p-rg-f-01-0-UI.s1 UI-R-C2p Rattus norvegicus cDNA clone
            UI-R-C2p-rg-f-01-0-UI 3', mRNA sequence.
ACCESSION   AI556365
NID         g4488728
VERSION     AI556365.1  GI:4488728
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 523)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     On May 18, 1998 this sequence version replaced gi:3137850.

            Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            Oligo-dT track not found, Not I site shown in beginning of sequence
            is likely internal to the message. cDNA Library Preparation: M.
            Fatima Bonaldo, Ph.D. Clone distribution: clones will be available
            through Research Genetics (www.resgen.com)
            Seq primer: M13 Forward.
FEATURES             Location/Qualifiers
     source          1. .523
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /map="MMU16C3-C4 region; MMU16C3-C4 region"
                     /clone="UI-R-C2p-rg-f-01-0-UI"
                     /clone_lib="UI-R-C2p"
                     /dev_stage="adult"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-C2p
                     library is a subtracted library derived from the UI-R-C1
                     library, which is a subtracted library derived from the
                     UI-R-C0 library.  The UI-R-C0 library consisted of a
                     mixture of individually tagged normalized  libraries
                     constructed from rat placenta, adult lung, brain, liver,
                     kidney,  heart, spleen, ovary, muscle, 8, 12 and 18-day
                     embryo. The tag is a string of  3-5 nucleotides present
                     between the Not I site and the oligo-dT track which allows
                     identification of the library of origin of a clone within
                     the mixture. The subtracted library (UI-R-C2p) was
                     constructed as follows: PCR amplified cDNA inserts from
                     UI-R-C1 clones from which 3' ESTs had been derived was
                     used as a driver in a hybridization with the UI-R-C1
                     library in the form of single-stranded circles. The
                     remaining single-stranded circles (subtracted library) was
                     purified by hydroxyapatite column chromatography,
                     converted to double-stranded circles and electroporated
                     into DH10B bacteria (Life Technologies) to generate the
                     UI-R-C2p library. This procedure has been previously
                     described (Bonaldo, Lennon and Soares, Genome Research 6:
                     791-806, 1996)"
BASE COUNT       96 a    112 c    192 g    121 t      2 others
ORIGIN


alignment_scores:
            Quality:   80.50              Length:    160
              Ratio:  0.904                Gaps:      4
 Percent Similarity:  55.625   Percent Identity: 20.000

alignment_block:
 US-09-068-804-7 x AI556365/rev   ..

 Align seg 1/1  to reverse of: AI556365  from: 1  to: 523

      6 AsnTyrSerAlaSerProHisProGly..................... 14
        ::::::::::|||:::::   ||||||
    510 TCGTTCCCAGCCACCAGCTCTCCAGGGGACTCCACCAGCACTGCAGTCAC 461

     15 .IleValAlaGluArgProGlnThrProSerAlaSerGluHisValGluT  31
           :::      |||   |||:::::::::::::::         |
    460 CAGTGGCACCTCGTCCCCAGCCACCAGCCCTCCAGGGGACTCCTCCAGCA 411

     31 hrAlaValValProSerThrThrGluHisArgGlyThrAspIleIleSer  47
        ||||||||::::::::|||:::            :::   |||::::::
    410 CTGCAGTCACCAGTGGCACCTCGTCCCCAGCCACCAGCCNTATATTGGAC 361

     47 ..................................................  47

    360 TCCTCCAGCACTGCAATCCCCAATGGTGCCTCGTCTCCGGCCACCAGCCN 311

     48 ....LeuSerGlnAlaAlaThrLysIleHisGlnAlaGlnGlnThrLeuG  63
           |||:::   ::::::|||   |||   ::::::      :::
    310 TATATTGGACTCCTCCAGCACTGCAATCCCCAATGGTGCCTCGTCTCCGG 261

     63 lnSerThrProProIleSerGluGluAsnAsnAspGluArgThrLeuAla  79
        ::::::||||||   :::      :::::::::      :::   |||
    260 CCACCAGCCCTCCAGGGGACTCCTCCAGCACTGCAGTCCTCAGCGGTGCT 211

     80 ArgGlnGlnLeuThrSerSerLeuAsnAlaLeuAla.....Ly  92
           |||   |||   ::::::::::::||||||
    210 TCCACCCAAACCACCAAGGCTGTATCAGACTTAGCTAGCACTCCAACCCA 161

     92 sSerGlyValSerLeuSerAlaGluGlnAsnGluAsnLeuArgSerAlaP 109
        ::::|||:::   :::::::::                   |||||||||:
    160 CAATGGCATCATGGTCCCAACTACC.............AGGTCTGCAT 126

    109 heSerAlaProThrSerAlaLeuPheSerAlaSerProMetAlaGlnPro 125
        ::|||::::::||||||::::::::|||:::|||:::::::::   |||
    125 TGAGCTCAGCCACCAGTCCGGTGTACAGTGGTTCTCAGAAGGACTGGCCT 76

    126 ArgThrThrIleSerAspAlaGluIleTrp 135
```

```
          |||:::::::   :::::::::::::::|||
       75 AGAAAACCTGGAGATGGGACCCAAATGTGG 46

seq_name: gb_est29:AI552661

seq_documentation_block:
LOCUS       AI552661      427 bp    mRNA            EST       23-MAR-1999
DEFINITION  ml98g09.y1 Stratagene mouse kidney (#937315) Mus musculus cDNA
            clone IMAGE:520096 5' similar to WP:C11D2.4 CE16845 ;, mRNA
            sequence.
ACCESSION   AI552661
NID         g4485024
VERSION     AI552661.1  GI:4485024
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 427)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
            Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
            Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     On Mar 10, 1998 this sequence version replaced gi:2949250.

            Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            This read is a RESEQUENCE of a previously sequenced mouse clone
            This read has been verified (found to hit its original self in the
            correct orientation)
            Seq primer: -40RP from Gibco
            High quality sequence stop: 410.
FEATURES             Location/Qualifiers
     source          1. .427
                     /organism="Mus musculus"
                     /strain="C57/Bl6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:520096"
                     /clone_lib="Stratagene mouse kidney (#937315)"
                     /sex="females"
                     /tissue_type="kidney"
                     /dev_stage="4 weeks"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Organ: kidney; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. Average insert size: 1.0 kb; Uni-ZAP XR Vector;
                     -5' adaptor sequence: 5' GAATTCGGCACGAG 3' -3' adaptor
                     sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT       67 a    137 c    153 g     69 t      1 others
ORIGIN


alignment_scores:
           Quality:   76.50            Length:    103
             Ratio:  1.159              Gaps:      5
 Percent Similarity:  64.078   Percent Identity: 25.243

alignment_block:
 US-09-068-804-7 x AI552661/rev   ..

 Align seg 1/1  to reverse of: AI552661  from: 1  to: 427

      31 ThrAlaValValProSerThrThrGluHisArgGlyThrAspIleIleSe 47
         :::::::|||      :::::::::::::||||||      ::::::::::::

     304 GCCAGTGTAGGGCGTCCCGCCGTAGCGCACCGCACACTTGCTGTCCTCCC 255

      47 rLeuSerGlnAlaAla......ThrLysIleHisGlnAlaGlnGlnThrL 62
         :|||:::::::::::::         ::::::||||||
     254 GCTCGGCCCAGAAGGAGAAGTTGAGCGAGTCCACCAG..... 218

      62 euGlnSerThrProProIleSerGluGluAsnAsnAspGluArgThrLeu 78
                                   |||:::   :::||||||:::|||
     217 ........................GAACACCCAGCCGAGCGCAGCCTC 194

      79 .........AlaArgGlnGlnLeuThrSerSerLeuAsnAlaLeu.AlaL 92
                 ||||||   :::|||:::   :::|||      ||| ::::
     193 GTCCGCCCCGCGCGGGTTGAGCTCGTGCAGGGACTTCCACTGCTCCACGC 144

      92 ysSerGlyValSerLeuSerAlaGluGlnAsnGluAsnLeuArg...Ser 107
         ::   :::::::::   ::::::      ::::::::::||||||   :::
     143 GCCACGCCGCGGCCGCGGGCAGCAGCAGCTCGGCCGCCCTGCGCACGCCC 94

     108 AlaPheSerAlaProThrSerAlaLeuPheSerAlaSerProMetAlaGl 124
         :::   |||::::::||||||   ||||||||||||:::   :::|||::
      93 TCGTGGTCCACGAGCACGTCGCGACTGTTCTCGGCCACGAACCTCGCCGA 44

     124 nProArg 126
         ::::|||
      43 CTCGCGC 37

seq_name: gb_est6:W01887

seq_documentation_block:
LOCUS       W01887        471 bp    mRNA            EST       18-APR-1996
DEFINITION  za35h04.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone
            IMAGE:294583 5', mRNA sequence.
ACCESSION   W01887
NID         g1273935
VERSION     W01887.1  GI:1273935
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 471)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     On Apr 14, 1993 this sequence version replaced gi:838025.

            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: mob.REGA+ET
            High quality sequence stop: 386.
FEATURES             Location/Qualifiers
     source          1. .471
                     /organism="Homo sapiens"
                     /db_xref="GDB:1239504"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:294583"
                     /clone_lib="Soares fetal liver spleen 1NFLS"
                     /sex="male"
                     /dev_stage="20 week-post conception fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                     with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                     1st strand cDNA was primed with a Pac I - oligo(dT) primer
```

[5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 135 a 120 c 111 g 105 t
ORIGIN

alignment_scores:
Quality: 76.00 Length: 82
Ratio: 1.310 Gaps: 3
Percent Similarity: 70.732 Percent Identity: 26.829

alignment_block:
US-09-068-804-7 x W01887 ..

Align seg 1/1 to: W01887 from: 1 to: 471

```
     57 GlnAlaGlnGlnThrLeuGlnSerThrProProIleSerGluGluAsnAs 73
        :::|||   |||   ::::::::::::::::|||||||||   ::::::::
     32 AGGGCCCTACAACATATTCCAGGCATATCTCCCATCTCCACTGCCCACCA 81

     73 nAspGluArgThrLeuAlaArgGln.................GlnLeuT 84
        :|||::::::|||||||:::|||:::                 :::|||:
     82 CGACCATCAAACATTAAGTAGGAATGGGCAAAGGTATGTTGCAAGACTAT 131

     84 hrSerSerLeuAsnAlaLeuAlaLys..............SerGlyVal 95
        ::::::::   :::|||:::   :::              |||::::::
    132 CTCCCTGCCACCAAGCTATGCAGAGAGCAGAAACAGAGAACAGTGATATA 181

     96 SerLeuSerAlaGluGlnAsnGluAsnLeuArg...SerAlaPheSerAl 111
           |||::::::   |||:::|||:::|||||   ::::::|||
    182 ATTCTCACCACCACTCAAAGTGAAACTTTGAGATGAGCAACTTTCAGATG 231

    111 aProThrSerAlaLeuPheSerAlaSerProMetAlaGlnProArg 126
         ::::::|||   |||:::::::::|||   ::::::   |||:::
    232 CTCATCTTCCTGTTTACTCCCTACGTCAAAAGTCAGTCTGCCCCAG 277
```

seq_name: gb_est26:AI324162

seq_documentation_block:
LOCUS AI324162 735 bp mRNA EST 23-DEC-1998
DEFINITION mh66a06.x1 Soares mouse placenta 4NbMP13.5 14.5 Mus musculus cDNA clone IMAGE:455890 3' similar to TR:Q99721 Q99721 TBP-ASSOCIATED FACTOR ;, mRNA sequence.
ACCESSION AI324162
NID g4058591
VERSION AI324162.1 GI:4058591
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 735)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Jan 14, 1998 this sequence version replaced gi:1798322.

Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:272778
This clone was previously sequenced on the 5' end only, this new data is from the 3' end
Possible reversed clone: similarity on wrong strand
High quality sequence stop: 459.

FEATURES Location/Qualifiers
source 1. .735
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:455890"
/clone_lib="Soares mouse placenta 4NbMP13.5 14.5"
/sex="unknown"
/tissue_type="placenta"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: placenta; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCGGAAATTTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 185 a 242 c 175 g 133 t
ORIGIN

alignment_scores:
Quality: 77.50 Length: 142
Ratio: 0.957 Gaps: 6
Percent Similarity: 57.042 Percent Identity: 25.352

alignment_block:
US-09-068-804-7 x AI324162 ..

Align seg 1/1 to: AI324162 from: 1 to: 735

```
      1 MetLeuAsnIleGlnAsnTyrSerAlaSerProHisProGlyIleValAl 17
        :::   ::::::|||:::::::::   ::::::   ::::::::::::||
    187 CTGGCCCAGATGCAGGCCCATGCCCAGGCCCAGCCTCAGAGTACCATGGC 236

     17 aGluArgProGlnThrProSerAlaSerGluHisValGluThrAlaValV 34
        |   ||||||   ||||||:::::::::      |||:::::::::::|
    237 ACCCCGTCCTGCAACCCCCACAGGTGCTCCACCTGTCCAGATCTCCACTG 286

     34 alProSerThrThrGluHisArgGlyThrAspIleIleSerLeuSerGln 50
        ||:::::::::           ||||||   ||||||:::
    287 TGCAGGCTCCA...........GGAACACCTATTATTGCACGGCAGGTG 324

     51 AlaAlaThrLysIle...........HisGlnAlaGlnGlnThrLeuGl 63
        ::::::|||:::|||           ||||||||   |||:::||
    325 ACCCCAACTACCATAATTAAGCAAGTGTCTCAGGCCCAGACGACGGTACA 374

     63 nSerThrProProIleSerGluGluAsnAsnAspGluArgThrLeuAlaA 80
        |:::|||:::                          ||||||   |
    375 GCCCACTACA.........................ACCCTACAGC 394

     80 rgGlnGlnLeuThrSerSerLeuAsnAlaLeuAlaLysSerGlyValSer 96
        ||   :::   :::   :::      :::|||:::   |||:::   :::
    395 GATCTCCCGGGGTACAGCCTCAGCTGGTTCTGGGTGGCTCCGCACAGCCA 444

     97 LeuSer.AlaGluGlnAsnGluAsnLeuArgSerAlaPheSerAlaProT 113
          |||   ||||||   :::   :::|||   :::
    445 GCCTCACTTGGAACAGCCACAGCTGTTCAGACAGGGA...CACCTCAACG 491

    113 hrSerAlaLeuPheSerAlaSerProMetAlaGlnProArgThrThrIle 129
          :::|||   :::|||:::|||:::      |||||||:::
    492 AACAGTCCCTGGGGCCAGCACCACCTCTACAGCTGCCACGGAAA..... 535
```

```
     130 SerAspAlaGluIleTrpAspMet 137
                 :::|||   |||
     536 ...........CTATGGAAAATG 547

seq_name: gb_est21:AA937937

seq_documentation_block:
LOCUS       AA937937      355 bp    mRNA            EST       13-MAY-1998
DEFINITION  oc08b01.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1340233 3',
            mRNA sequence.
ACCESSION   AA937937
NID         g3096048
VERSION     AA937937.1  GI:3096048
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 355)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2152037.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
            Ph.D., Gerald Marti, M.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 2003   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 291.
FEATURES             Location/Qualifiers
     source          1. .355
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="22"
                     /clone="IMAGE:1340233"
                     /clone_lib="NCI_CGAP_GCB1"
                     /tissue_type="germinal center B cell"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from human tonsillar cells enriched for
                     germinal center B cells by flow sorting (CD20+, IgD-),
                     provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                     (NCI) and Dr. Gerald Marti (CBER).  cDNA synthesis was
                     primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTT-
                     3']. Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT       76 a     95 c    116 g     67 t      1 others
ORIGIN


alignment_scores:
           Quality:   73.50            Length:    128
             Ratio:  0.930              Gaps:      6
 Percent Similarity:  61.719   Percent Identity: 24.219

alignment_block:

US-09-068-804-7 x AA937937   ..

Align seg 1/1  to: AA937937  from: 1  to: 355

     18 GluArgProGlnThrProSerAlaSerGluHisValGluThrAlaValVa 34
        |||:::|||||:::::::::|||:::   |||         :::|||
      2 GAACAGCCACAAATCTCATGCGCCTGC...CACTCAGTGTTNGGGGTAAG 48

     34 lProSerThrThrGluHisArgGlyThrAspIleIleSerLeuSerGlnA 51
        ||||||   :::   :::|||||||         :::::         :
     49 TCCTAGCAGAGCTTCTGAGCGTGGT......TATTTAGGAACCATGAGTC 92

     51 laAlaThrLysIleHisGlnAlaGlnGlnThrLeuGlnSerThrProPro 67
        ::|||   :::   :::   :::         |||||:::   |||:::
     93 CAGCTAGGATGGGGCAGGCCACCTCCTGGGGGCTGCAGGCAGAACCTCAG 142

     68 IleSerGluGluAsnAsnAspGluArgThrLeuAlaArgGlnGlnLeuTh 84
           |||:::::::::   :::   ::::::   |||
    143 GAGAGGACAGAGAAAGGAGGTGGGAAGCAAGTCACGATTCAG......CT 186

     84 rSer.SerLeuAsnAlaLeuAlaLysSerGlyValSerLeuSerAlaGlu 100
        ::: :::::::::   |||:::::::::||||||:::   :::|||:::
    187 CACACGGTGTAGACAGGCTGGGCAGAGGCGGGGTGGACCACGCTGCAGAC 236

    101 GlnAsnGluAsnLeuArgSerAlaPheSerAlaProThrSerAlaLeuPh 117
        |||::::::                 :::::|||   ||||||:::::
    237 CAGCAGGAT..............GTGGCTGTCCCTGGCAGTGCCATGCT 271

    117 eSerAlaSerProMetAlaGlnProArgThrThrIleSerAspAlaGluI 134
        :         |||::::::      |||:::::|||:::         |
    272 G.........CCTCTTTCTGCCTTCCGGGCCAGCATCACCACACGGCCCA 312

    134 leTrpAspMetValSerGlnAsnIleSerAla 144
        |||||   ||||||:::   ::::::::::::
    313 TATGGATGATGGTTACTTTGGCAGTCGGCCCT 344

seq_name: gb_est27:AI398188

seq_documentation_block:
LOCUS       AI398188      506 bp    mRNA            EST       08-FEB-1999
DEFINITION  NCSM2F10T3 Subtracted Mycelial Neurospora crassa cDNA clone SM2F10
            5', mRNA sequence.
ACCESSION   AI398188
NID         g4241273
VERSION     AI398188.1  GI:4241273
KEYWORDS    EST.
SOURCE      Neurospora crassa.
  ORGANISM  Neurospora crassa
            Eukaryota; Fungi; Ascomycota; Euascomycetes; Pyrenomycetes;
            Sordariales; Sordariaceae; Neurospora.
REFERENCE   1  (bases 1 to 506)
  AUTHORS   Nelson,M.A., Kang,S., Braun,E.L., Crawford,M.E., Dolan,P.L.,
            Leonard,P.M., Mitchell,J., Armijo,A.M., Bean,L., Blueyes,E.,
            Cushing,T., Errett,A., Fleharty,M., Gorman,M., Judson,K.,
            Miller,R., Ortega,J., Pavlova,I., Perea,J., Todisco,S.,
            Trujillo,R., Valentine,J., Wells,A., Werner-Washburne,M., Yazzie,S.
            and Natvig,D.O.
  TITLE     Expressed sequences from conidial, mycelial, and sexual stages of
            Neurospora crassa
  JOURNAL   Fungal Genet. Biol. 21, 348-363 (1997)
  MEDLINE   97435549
COMMENT     On Mar 10, 1998 this sequence version replaced gi:2948382.

            Contact: Natvig,D.O./Nelson,M.A.
            Department of Biology
            University of New Mexico
            Castetter Hall, Albuquerque, NM 87131, USA
            Tel: 505 277 3411
            Fax: 505 277 0304
            Email: ngp@biology.unm.edu.
FEATURES             Location/Qualifiers
     source          1. .506
                     /organism="Neurospora crassa"
```

```
                     /strain="74-OR23-IV A (FGSC 2489)"
                     /db_xref="taxon:5141"
                     /map="8q24.11"
                     /clone="SM2F10"
                     /clone_lib="Subtracted Mycelial"
                     /sex="Mating type A"
                     /tissue_type="Mycelium"
                     /dev_stage="Mycelium"
                     /lab_host="E. coli"
                     /note="Vector: pBlueScript SK (-); Site_1: EcoRI; Site_2:
                     XhoI; 2% sucrose for 24 hours.  cDNA directionally cloned
                     into pBluescript  SK(-) using the Uni-ZAP XR vector system
                     (Stratagene, La Jolla,  CA). Previously identified highly
                     expressed clones were subtracted  from this library."
BASE COUNT      118 a    181 c    118 g     89 t
ORIGIN


alignment_scores:
            Quality:   75.00             Length:    129
              Ratio:  1.014               Gaps:      5
 Percent Similarity:  57.364   Percent Identity: 18.605

alignment_block:
 US-09-068-804-7 x AI398188   ..

 Align seg 1/1  to: AI398188  from: 1  to: 506

       1 MetLeuAsnIleGlnAsn......TyrSerAlaSerProHisProGlyIl 15
         |||:::::::|||   :::      |||:::::::|||:::   |||:::::
     140 ATGATGTCTATCAGCCAGACTCCATATCCAGTTTCCGCTGGCCCCTCAAC 189

      15 eValAlaGluArgProGlnThrProSerAlaSerGluHisValGluThrA 32
         :   |||      |||   ||||||:::   :::      :::::::|||:
     190 CGGTGCCTCGGCCCCGACAACACCAGGCCTTGGT...CCCACTCACACCG 236

      32 laValValProSerThrThrGluHisArgGlyThrAspIleIleSerLeu 48
         ::::::::::   :::::::::   :::|||::::::   :::   :::
     237 GGCTGTTGGCTTTGCCGCCCCAACCTCAAGGAGTTCAGAACGTTCAGGTG 286

      49 SerGlnAlaAlaThrLysIleHisGlnAlaGlnGlnThrLeuGlnSerTh 65
         :::::::::            |||||||:::   :::|||::::::::::
     287 GCACACGGCCAT.........CACCAAGTGATGCACACCTTCCCCCCGCT 327

      65 rProPro......IleSerGluGluAsnAsnAspGluArgThrLeuAlaA 80
          ||||||      ::::::   ::::::         ::::::::::::
     328 TCCACCACATATCATGCCTCCGCAGCATATGCCTCCGCAATCGATGCCCC 377

      80 rgGlnGlnLeuThrSerSerLeuAsnAlaLeuAlaLysSerGlyValSer 96
           ||||||::::::::::
     378 CTCAGCAAATGCCACCT.............................. 394

      97 LeuSerAlaGluGlnAsnGluAsnLeuArgSerAlaPheSerAlaProTh 113
                                 ||||||      :::   |||
     395 ........................CGCTCAATGGCTCCTCAACCCAT 417

     113 rSerAlaLeuPheSerAlaSerProMetAlaGlnPro 125
          :::::::::          :::||||||::::::|||
     418 GCCTCCAATGGCCCAGATGGCTCCCATGTCTCATCCT 454

 seq_name: gb_est16:AA550027

 seq_documentation_block:
 LOCUS       AA550027      593 bp    DNA              EST       11-AUG-1997
 DEFINITION  1106m3 gmbPfHB3.1, G. Roman Reddy Plasmodium falciparum genomic
             clone 1106m, mRNA sequence.
 ACCESSION   AA550027
 NID         g2320279
 VERSION     AA550027.1  GI:2320279
 KEYWORDS    EST.
 SOURCE      malaria parasite P. falciparum.
   ORGANISM  Plasmodium falciparum
             Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 REFERENCE   1  (bases 1 to 593)
   AUTHORS   Dame,J.B., Arnot,D.E., Bourke,P., Chakrabarti,D., Christodoulou,Z.,
             Coppel,R., Cowman,A., Craig,A., Fischer,K., Foster,J., Goodman,N.,
             Hinterberg,K., Holder,A.A., Holt,D., Kemp,D., Lanzer,M., Lim,A.,
             Newbold,C., Ravetch,J.V., Reddy,G.R., Rubio,J., Schuster,S.M.,
             Su,X.-Z., Thompson,J.K., Vital,F., Wellems,T.E. and Werner,E.
   TITLE     Current status of the Plasmodium falciparum genome project
   JOURNAL   Mol. Biochem. Parasitol. 79, 1-12 (1996)
   MEDLINE   97001675
 COMMENT     On May 8, 1995 this sequence version replaced gi:800989.

             Contact: Dame JB
             Department of Pathobiology
             University of Florida
             2015 SW 16th Ave., Bldg. 1017, Gainesville, FL  32611-0880
             Tel: 352 392 4700
             Fax: 352 392 9704
             Email: dame@icbr.ifas.ufl.edu
             Seq primer: T3.
 FEATURES             Location/Qualifiers
      source          1. .593
                      /organism="Plasmodium falciparum"
                      /db_xref="taxon:5833"
                      /clone="1106m"
                      /clone_lib="gmbPfHB3.1, G. Roman Reddy"
                      /lab_host="E. coli XL1-Blue"
                      /note="Vector: pBlueScript SK(+); Genomic DNA, from
                      asynchronous blood stage parasites of the cloned Honduran
                      HB3 isolate cultured in vitro, was digested with mung bean
                      nuclease in the presence of 30% formamide at 50oC
                      (Vernick, K.D., Imberski, R.B., and McCutchan, T.F. 1988.
                      Nucleic Acids Research 16:6883-6896).  The ends of the
                      fragments were polished using T4 DNA polymerase, and the
                      fragments were ligated to EcoR V-cleaved and
                      dephosphorylated pBlueScript SK(+).  Recombinant plasmids
                      transformed E. coli XL1-Blue."
 BASE COUNT      150 a     89 c     67 g    286 t      1 others
 ORIGIN


 alignment_scores:
             Quality:   75.50             Length:    196
               Ratio:  0.624               Gaps:      9
  Percent Similarity:  61.735   Percent Identity: 17.857

 alignment_block:
  US-09-068-804-7 x AA550027/rev   ..

  Align seg 1/1  to reverse of: AA550027  from: 1  to: 593

      16 ValAlaGluArgProGlnThrProSerAlaSerGluHisValGluThrAl 32
         |||:::      |||   :::|||   ::::::::::::   |||:::::
     542 GTACCACCTCCTCCAGCACCACCCTTAACAAATAATCAAGGTGAAAATAA 493

      32 aValValProSerThrThrGluHisArgGlyThrAspIleIleSerLeuS 49
         ::::         :::|||::::::|||:::::::::          |||
     492 TATA.........CCTACAGATAATAGAAATAACGAAAATAAAAAATTGA 452

      49 erGlnAlaAlaThrLysIleHisGlnAlaGlnGlnThrLeuGlnSerThr 65
           ::::::::::::|||::::::|||
     451 TGAATACAAATATAAAAATGAATCAA..................... 426

      66 ProProIleSerGluGluAsnAsnAspGluArgThrLeuAlaArgGlnGl 82
           ::::::::::::::::||||||::::::   :::
     425 ...CAAGTTGATAACAATAATAATAATAATAATAATAATTTGGGTAGTGC 379

      82 nLeuThrSerSerLeuAsnAlaLeuAlaLysSerGlyValSerLeuSerA 99
         ::::::::::::::|||::::::::::|||::::||          ::::
     378 AATGAATACACCTATGAATAATATGAACAAGGGGGGCCCAAGAAACAATG 329

      99 laGluGlnAsnGluAsnLeuArgSerAlaPheSerAlaProThrSerAla 115
         ::::::::::|||::::::|||   :::   |||:::   |||   :::
     328 TGAACAATTCTGAACATATGAGG...GAACATAGTAATCATACGGTAAAT 282
```

```
     116 LeuPheSerAlaSerProMetAlaGlnProArgThrThrIleSerAspAl 132
         |||      :::||||||:::|||                 ::::::::
     281 AGATTT......GCGCCGATGGGACAA............TATGCAAACAG 250

     132 aGluIleTrpAspMetValSer...GlnAsnIleSerAlaIleGlyAspS 148
         :          :::|||   :::   :::|||:::::::::::::::::::
     249 CAAAAATAATAATATGAATAATGTGCATAATGTGAATAATGTGAATAATG 200

     148 erTyrLeuGlyValTyrGluAsnValValAlaValTyrThrAspPheTyr 164
                     :::||||||   :::|||
     199 TG..............AATAATGTGAATAATGTG.............. 180

     165 GlnAlaPheSerAspIleLeuSerLysMetGlyGlyTrpLeuLeuProGl 181
                   :::   :::::::::|||
     179 ............AATAATATGAACAATATG..................... 162

     181 yLysAspGlyAsnThrValLysLeuAspValThrSerLeuLysAsnAspL 198
          ::::::   |||:::::::::   :::::::::::::::::::|||::::
     161 .AACAATATGAACAATATGAACAATAATATGAACAATATGAACAATAATA 113

     198 euAsnSerLeuValAsnLysTyrAsnGlnIleAsnSer 210
         ::|||:::   :::|||:::   |||::::::||||||
     112 TGAACAATAATATGAACAATATGAACAATATGAATAGT 75

seq_name: gb_est27:AI456803

seq_documentation_block:
LOCUS       AI456803      509 bp    mRNA            EST       09-MAR-1999
DEFINITION  LD36936.5prime LD Drosophila melanogaster embryo pOT2 Drosophila
            melanogaster cDNA clone LD36936 5prime, mRNA sequence.
ACCESSION   AI456803
NID         g4283215
VERSION     AI456803.1  GI:4283215
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 509)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     On May 18, 1998 this sequence version replaced gi:3137536.

            Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            hit genomic sequence AC006092
            Plate: 369  row: C  column: 12
            High quality sequence stop: 508.
FEATURES             Location/Qualifiers
     source          1. .509
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /map="9"
                     /clone="LD36936"
                     /clone_lib="LD Drosophila melanogaster embryo pOT2"
                     /sex="male and female"
                     /dev_stage="0 to 24 hours mixed stage embryonic"
                     /lab_host="XL1 Blue"
                     /note="Organ: embryo; Vector: pOT2; Site_1: EcoRI; Site_2:
                     XhoI; Sized fractionated cDNAs were directly ligated into
                     pOT2. "
BASE COUNT      148 a    150 c    122 g     89 t
ORIGIN

alignment_scores:
            Quality:   73.50            Length:    104
              Ratio:  1.131              Gaps:      3
 Percent Similarity:  62.500   Percent Identity: 24.038

alignment_block:
 US-09-068-804-7 x AI456803   ..

 Align seg 1/1  to: AI456803  from: 1  to: 509

      10 SerProHisProGlyIleValAlaGluArgProGlnThrProSerAlaSe 26
         ||||||   |||      |||::::::|||      ||||||   :::::
     207 AGCCCGGCCCCTCCTGCTGTGCCAGACCGC...GTAACTCCACATTCAAC 253

      26 rGluHisValGluThrAlaValValProSerThrThrGluHisArgGlyT 43
         :|||         :::::::::   :::::::::|||      |||::::
     254 GGAAATTTCAGTTGCCCCCGCCAATTCTACAAGCACAACAGTACGAGCAG 303

      43 hrAspIleIleSerLeuSerGlnAlaAlaThrLysIleHisGlnAlaGln 59
         :::::   ::::::   :::   :::|||||||:::   |||:::   :::
     304 CAGGATCAGTAGGAGCAGCCTTGCCGGCCACCCGC...CATCACCAACAT 350

      60 GlnThrLeuGlnSerThrProProIleSerGluGluAsnAsnAspGluAr 76
         :::   |||   :::       |||          ::::::::::::::
     351 ATAGCGACCCAAGTGAAGGGAATCGCCAGCAGCAGCAGCAAACAACAGAA 400

      76 gThrLeuAlaArgGlnGlnLeuThrSerSerLeuAsnAlaLeuAlaLysS 93
         :   ||||||      ||||||:::   :::|||::::::|||::::::
     401 GCAACTGGCCAGTGCGCAGCTGCCTGTGCCGCTGTCGCCCITGCCGCAA. 449

      93 erGlyValSerLeuSerAlaGluGlnAsnGluAsnLeuArgSerAlaPhe 109
                              :::|||:::::::::      :::|||
     450 .............CAACAACAGCAAACGGCAGAGGCAACGGCAGCAGCA 485

     110 SerAlaProThr 113
         :::|||||||:::
     486 GCAGCACCCGCC 497

seq_name: gb_est28:AI488463

seq_documentation_block:
LOCUS       AI488463      492 bp    mRNA            EST       12-APR-1999
DEFINITION  EST246802 tomato ovary, TAMU Lycopersicon esculentum cDNA clone
            cLED21B2, mRNA sequence.
ACCESSION   AI488463
NID         g4383834
VERSION     AI488463.1  GI:4383834
KEYWORDS    EST.
SOURCE      tomato.
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
            Potatoe; Lycopersicon.
REFERENCE   1  (bases 1 to 492)
  AUTHORS   Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T.,
            Chandra,I., Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman,C.L.,
            Nierman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley,S.D.
            and Giovannoni,J.
  TITLE     Generation of ESTs from tomato carpel tissue
  JOURNAL   Unpublished (1999)
COMMENT     On Jan 24, 1997 this sequence version replaced gi:1798256.

            Contact: David Frisch
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 4366
            Fax: 864 656 4293
            Email: dfrisch@CLEMSON.EDU.
FEATURES             Location/Qualifiers
     source          1. .492
                     /organism="Lycopersicon esculentum"
```

```
                     /cultivar="TA496"
                     /db_xref="taxon:4081"
                     /map="12; 672E12; 17; 17q23-q25"
                     /clone="cLED21B2"
                     /clone_lib="tomato ovary, TAMU"
                     /tissue_type="carpel"
                     /dev_stage="5 days pre-anthesis to 5 days post-anthesis"
                     /lab_host="XL1-Blue MRF'"
                     /note="Vector: pBlueScript SK(-); Site_1: EcoR1; Site_2:
                     Xho1; cLED - Tomato Carpel EST Library. OligodT-primed and
                     directionally cloned  cDNA in vector Lamda ZAP II with 5'
                     and 3' ends located at the EcoRI and  XhoI sites,
                     respectively."
BASE COUNT      135 a    140 c     88 g    129 t
ORIGIN


alignment_scores:
           Quality:   73.00            Length:    105
             Ratio:  1.058              Gaps:      6
 Percent Similarity:  65.714   Percent Identity: 22.857

alignment_block:
 US-09-068-804-7 x AI488463   ..

 Align seg 1/1  to: AI488463  from: 1  to: 492

      18 GluArgProGlnThrProSerAlaSerGluHisValGluThrAlaValVa 34
         :::   ||||||   |||   :::::::::||||||      ::::::||
      23 CAGCCCCCTCAGCTTCCTCGTCCTCCACAACATGTT...AGGCCAATTGT 69

      34 lProSerThrThrGluHisArgGlyThrAspIleIleSerLeuSerGlnA 51
         ||||:::::::::::::      :::   :::   ::::::|||   |||:
      70 TCCAGCTTCGCCCCAG......TCAGAGCAAAGTGTTCCTTTGCTACAAA 113

      51 laAlaThrLysIleHis...GlnAlaGlnGlnThrLeuGlnSerThrPro 66
         :::::      ::::::   |||::::::|||:::|||:::      |||
     114 GCCCCATGCACATGCAAATGCAATCACCACAGATCTTGCAT...CAACCT 160

      67 ProIleSer.................GluGluAsnAsnAspGluArgTh 77
         :::::::|||                :::   ::::::::|||   ::
     161 CAAGTTTCTCCAGCTCATGTGTACTACCAGACTCAACAGCAGGAGAATTC 210

      77 rLeuAlaArgGlnGlnLeuThrSerSerLeuAsnAlaLeuAlaLysSerG 94
         :|||   :::||||||:::      ||||||:::   :::::::::
     211 GTTACAACAGCAACAAATCGAACATTCTCTTTCTCAAGTACCACAGCAGC 260

      94 lyValSerLeuSerAlaGluGlnAsnGlu.........AsnLeuArgSer 107
            ::::::   ::::::|||::::::        :::|||::::::
     261 AAGGTGATATTGTCACCCAGCAACAAGATTCTGGGATGTCATTACAAGAT 310

     108 AlaPheSerAlaPro 112
            |||   :::|||
     311 TTTTTTAGATCTCCA 325

seq_name: gb_est19:AA777598

seq_documentation_block:
LOCUS       AA777598      404 bp    mRNA            EST       05-FEB-1998
DEFINITION  zj03a10.s1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA
            clone IMAGE:449178 3' similar to TR:O00268 O00268 SUBUNIT OF RNA
            POLYMERASE II TRANSCRIPTION FACTOR TFIID. ;, mRNA sequence.
ACCESSION   AA777598
NID         g2837077
VERSION     AA777598.1  GI:2837077
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 404)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
            Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
            Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2152288.

            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Possible reversed clone: similarity on wrong strand
            Possible reversed clone: polyT not found
            Seq primer: -40m13 fwd. ET from Amersham.
FEATURES             Location/Qualifiers
     source          1. .404
                     /organism="Homo sapiens"
                     /db_xref="GDB:1353435"
                     /db_xref="taxon:9606"
                     /map="16 q12.1"
                     /clone="IMAGE:449178"
                     /clone_lib="Soares_fetal_liver_spleen_1NFLS_S1"
                     /sex="male"
                     /dev_stage="20 week-post conception fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                     with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                     This is a subtracted version of the original Soares fetal
                     liver spleen 1NFLS library.   1st strand cDNA was primed
                     with a Pac I - oligo(dT) primer [5'
                     AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Pac I and cloned into the Pac I
                     and Eco RI sites of the modified pT7T3 vector.  Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT       84 a    151 c    103 g     66 t
ORIGIN


alignment_scores:
           Quality:   72.00            Length:     72
             Ratio:  1.412              Gaps:       2
 Percent Similarity:  70.833   Percent Identity: 27.778

alignment_block:
 US-09-068-804-7 x AA777598   ..

 Align seg 1/1  to: AA777598  from: 1  to: 404

       1 MetLeuAsnIleGlnAsnTyrSerAlaSerProHisProGlyIleValAl 17
         :::   :::::::|||:::::::::   ::::::   ::::::::::::||
     157 CTGGCCAGCATGCAGGCCCATGCCCAGGCCCAGCCTCAGAGTACCATGGC 206

      17 aGluArgProGlnThrProSerAlaSerGluHisValGluThrAlaValV 34
         |   ||||||   ||||||:::::::::       |||:::::::::::|
     207 ACCCCGTCCTGCAACCCCCACAGGTGCTCCACCTGTCCAGATCTCCACTG 256

      34 alProSerThrThrGluHisArgGlyThrAspIleIleSerLeuSerGln 50
         ||:::::::::              ||||||   |||||||:::
     257 TGCAGGCTCCA............GGAACACCTATTATTGCACGGCAGGTG 294

      51 AlaAlaThrLysIle............HisGlnAlaGlnGlnThrLeuGl 63
         ::::::|||:::|||               |||||||||   |||:::||
     295 ACCCCAACTACCATAATTAAGCAAGTGTCTCAGGCCCAGACGACGGTACA 344

      63 nSerThrProProIle 68
         |:::|||:::::::::
     345 GCCCACTACAACCCTA 360
```

```
seq_name: gb_est28:AI483429

seq_documentation_block:
LOCUS       AI483429      501 bp    mRNA            EST       12-APR-1999
DEFINITION  EST249250 tomato ovary, TAMU Lycopersicon esculentum cDNA clone
            cLED24A22, mRNA sequence.
ACCESSION   AI483429
NID         g4387353
VERSION     AI483429.1  GI:4387353
KEYWORDS    EST.
SOURCE      tomato.
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
            Potatoe; Lycopersicon.
REFERENCE   1  (bases 1 to 501)
  AUTHORS   Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T.,
            Chandra,I., Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman,C.L.,
            Nierman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley,S.D.
            and Giovannoni,J.
  TITLE     Generation of ESTs from tomato carpel tissue
  JOURNAL   Unpublished (1999)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3189532.

            Contact: David Frisch
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 4366
            Fax: 864 656 4293
            Email: dfrisch@CLEMSON.EDU.
FEATURES             Location/Qualifiers
     source          1. .501
                     /organism="Lycopersicon esculentum"
                     /cultivar="TA496"
                     /db_xref="taxon:4081"
                     /clone="cLED24A22"
                     /clone_lib="tomato ovary, TAMU"
                     /tissue_type="carpel"
                     /dev_stage="5 days pre-anthesis to 5 days post-anthesis"
                     /lab_host="XL1-Blue MRF'"
                     /note="Vector: pBlueScript SK(-); Site_1: EcoR1; Site_2:
                     Xho1; cLED - Tomato Carpel EST Library. OligodT-primed and
                     directionally cloned  cDNA in vector Lamda ZAP II with 5'
                     and 3' ends located at the EcoRI and  XhoI sites,
                     respectively."
BASE COUNT      140 a    138 c     91 g    132 t
ORIGIN


alignment_scores:
            Quality:   73.00            Length:    105
              Ratio:  1.058              Gaps:      6
 Percent Similarity:  65.714   Percent Identity: 22.857

alignment_block:
 US-09-068-804-7 x AI483429   ..

 Align seg 1/1  to: AI483429  from: 1  to: 501

      18 GluArgProGlnThrProSerAlaSerGluHisValGluThrAlaValVa 34
         :::   ||||||   |||   ::::::::::||||||      :::::::||
      23 CAGCCCCCTCAGCTTCCTCGTCCTCCACAACATGTT...AGGCCAATTGT 69

      34 lProSerThrThrGluHisArgGlyThrAspIleIleSerLeuSerGlnA 51
         ||||:::::::::::::      :::   :::   :::::::|||   |||:
      70 TCCAGCTTCGCCCCAG......TCAGAGCAAAGTGTTCCTTTGCTACAAA 113

      51 laAlaThrLysIleHis...GlnAlaGlnGlnThrLeuGlnSerThrPro 66
         :::::      ::::::   |||:::::::|||:::|||:::      |||
     114 GCCCCATGCACATGCAAATGCAATCACCACAGATCTTGCAT...CAACCT 160

      67 ProIleSer.................GluGluAsnAsnAspGluArgTh 77
         :::::::|||                  :::   :::::::::|||   ::
     161 CAAGTTTCTCCAGCTCATGTGTACTACCAGACTCAACAGCAGGAGAATTC 210

      77 rLeuAlaArgGlnGlnLeuThrSerSerLeuAsnAlaLeuAlaLysSerG 94
         :|||   :::|||||||:::       ||||||:::   :::::::::
     211 GTTACAACAGCAACAAATCGAACATTCTCTTTCTCAAGTACCACAGCAGC 260

      94 lyValSerLeuSerAlaGluGlnAsnGlu.........AsnLeuArgSer 107
            ::::::   :::::::|||:::::::         :::|||::::::
     261 AAGGTGATATTGTCACCCAGCAACAAGATTCTGGGATGTCATTACAAGAT 310

     108 AlaPheSerAlaPro 112
            |||   :::|||
     311 TTTTTTAGATCTCCA 325

seq_name: gb_est28:AI532519

seq_documentation_block:
LOCUS       AI532519      669 bp    mRNA            EST       18-MAR-1999
DEFINITION  SD04057.5prime SD Drosophila melanogaster Schneider L2 cell culture
            pOT2 Drosophila melanogaster cDNA clone SD04057 5prime, mRNA
            sequence.
ACCESSION   AI532519
NID         g4446654
VERSION     AI532519.1  GI:4446654
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 669)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     On Mar 10, 1998 this sequence version replaced gi:2949102.

            Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 40  row: E  column: 9
            High quality sequence stop: 502.
FEATURES             Location/Qualifiers
     source          1. .669
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /map="6; 6, 2.94cR from D6S442"
                     /clone="SD04057"
                     /clone_lib="SD Drosophila melanogaster Schneider L2 cell
                     culture pOT2"
                     /lab_host="DH5-alpha"
                     /note="Vector: pOT2; Site_1: EcoRI; Site_2: Xho1; Sized
                     fractionated cDNAs were directly ligated into pOT2.
                     Plasmid cDNA library."
BASE COUNT      196 a    230 c    141 g    102 t
ORIGIN


alignment_scores:
            Quality:   74.00            Length:    131
              Ratio:  0.949              Gaps:      5
 Percent Similarity:  59.542   Percent Identity: 22.901

alignment_block:
 US-09-068-804-7 x AI532519   ..

 Align seg 1/1  to: AI532519  from: 1  to: 669
```

```
      5 GlnAsnTyrSerAlaSerProHisProGlyIleValAlaGluArgProGl 21
        ::::::::::|||::::::|||   |||      |||::::::   :::::
    235 CACTCTCATTCGCCCACACCCTCGCCGCCC...GTCCCCGCGTCCTCTCA 281

     21 nThrProSerAlaSerGluHisValGluThrAlaValValProSerThrT 38
        :   ::::::   :::::::|||      :::::::::   :::|||   |
    282 CGACCAACCCCATCCCGATCACCCTAGTCCCCCATTGAATGCATCTGAAA 331

     38 hrGluHisArgGlyThrAspIleIleSerLeuSerGlnAlaAlaThr... 53
        |||||   |||   ::::::       :::   |||::::::::::|||
    332 CGGAAGCTCGCGAGAATTCACCACATGATCACTCACCCACTCCTACTTTC 381

     54 .............................................LysIl 55
                                                       :::
    382 CACCAGACTGCACCGCCACCCACGACTTCATCGACGGCACCGCAGCGGGA 431

     55 eHisGlnAlaGlnGlnThrLeuGlnSerThrProProIleSerGluGluA 72
         :::::::::|||||      |||   :::||||||:::   :::::::
    432 CGAGCGGGAGCAGCAGCAGCAGCAAGAGGCGCCACCATTGCAGCAGGATC 481

     72 snAsnAspGluArgThrLeuAlaArgGlnGlnLeuThrSerSerLeuAsn 88
        ::::::::::   :::        |||:::|||   :::   |||:::
    482 AGCAGGAGAATTCGCCCGCGCAGGACCAGGAGCTGCACCCATTGCTGGAT 531

     89 AlaLeuAlaLys......SerGlyVal......SerLeuSerAlaGluGl 101
           ::::::      :::::::|||      :::      ||||||||
    532 CAACAGAATCAAGAATATCCCGCAGTGCACCAAGATCAACAAGCGGAGCA 581

    101 nAsnGluAsnLeu.........ArgSerAlaPheSerAlaPro 112
        ||||::::::|||         |||:::|||:::|||||||||
    582 AAACCAGGAGCTGCATCACATTAGAGGGGCTTATTCGGCACCG 624
```

seq_name: gb_est1:T01512

```
seq_documentation_block:
LOCUS       T01512         331 bp    mRNA            EST       10-NOV-1992
DEFINITION  wEST02233 Early embryo, Stratagene (cat. #937007) Caenorhabditis
            elegans cDNA clone CEESR74, mRNA sequence.
ACCESSION   T01512
NID         g277993
VERSION     T01512.1  GI:277993
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 331)
  AUTHORS   McCombie,W.R., Kelley,J.M., Aubin,L., Goscoechea,M.,
            FitzGerald,M.G., Wu,A., Adams,M.D., Dubnick,M., Kerlavage,A.R.,
            Venter,J.C. and Fields,C.A.
  TITLE     Caenorhabditis elegans cDNAs
  JOURNAL   Unpublished (1993)
COMMENT
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            Seq primer: M13 Foward.
FEATURES             Location/Qualifiers
     source          1. .331
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /clone="CEESR74"
                     /clone_lib="Early embryo, Stratagene (cat. #937007)"
BASE COUNT      110 a     55 c     90 g     73 t      3 others
ORIGIN


alignment_scores:
            Quality:   70.50            Length:    139
                Ratio:  0.916                Gaps:      6
 Percent Similarity:  55.396   Percent Identity: 23.022

alignment_block:
 US-09-068-804-7 x T01512   ..

 Align seg 1/1  to: T01512  from: 1  to: 331

    177 TrpLeuLeuPro.............GlyLysAspGlyAsnThrValLysLe 189
        |||   |||:::                |||   ::::::|||:::
      1 TGGAAATTGGCAGTTTCGGCGGTCGAAAAGAAGAATAAAACGTTGATTGG 50

    189 uAspValThrSerLeuLysAsnAspLeuAsnSerLeuValAsnLysTyrA 206
         :::::::::   ||||||||||:::|||:::|||       |||   |||
     51 TGAGCTCAATGAGCTGAAGAATCAATTGGAGAGC...CGGAATGATTATN 97

    206 snGlnIleAsnSerAsnThrValLeuPheProAlaGlnSerGlySerGly 222
            |||::::::::::       |||:::   :::         |||
     98 AGGCTATTAAGAATGAATTGAGACTTCTTCGC...GAAATCGAATTCGGA 144

    223 ValLysValAlaThrGluAlaGluAlaArgGlnTrpLeuSerGluLeuAs 239
            :::::::::::::||||||:::   ::::::|||:::|||
    145 GATTCTGCCGAAGCGAACGCCGAATCAATTGAACGACTCGGTGAA..... 189

    239 nLeuProAsnSerCysLeuLysSerTyrGlySerGlyTyrValValThrV 256
                                                       ||||
    190 .............................................ACTG 193

    256 alAspLeuThrProLeuGlnLysMetValGlnAspIleAspGlyLeuGly 272
        ||:::      :::|||:::::::::::::   :::
    194 TGGAG......ACGCTGGATCGCCTGCTCGCCGAG.............. 222

    273 AlaProGlyLysAspSerLysLeuGluMetAspAsnAlaLysTyrGlnAl 289
                |||:::   :::|||:::   :::||||||      :::::
    223 .........AAAAATCGCCGTCTTCAGAATGAAAATGCTAGTCTTCGAGT 263

    289 aTrpGlnSerGlyPheLysAlaGlnGluGluAsnMetLysThrThrLeuG 306
        :   ::::::||||||||:::::::::|||:::      :::   |||
    264 TGCAAATNATGGATTTAAAGGACGNAATGAGGAGCAAGAAGCCGAATTGA 313

    306 lnThrLeuThrGlnLys 311
          :::|||::::::|||
    314 CAGTGCTCAAAGAGAAA 330
```

seq_name: em_est22:HSM007904

```
seq_documentation_block:
ID   HSM007904  standard; RNA; EST; 796 BP.
XX
AC   AL043054;
XX
SV   AL043054.1
XX
NI   e1401395
XX
DT   12-MAR-1999 (Rel. 59, Created)
DT   12-MAR-1999 (Rel. 59, Last updated, Version 1)
XX
DE   Homo sapiens mRNA; EST DKFZp434A2223_r1  (from clone
DE   DKFZp434A2223)
XX
KW   EST; expressed sequence tag.
XX
OS   Homo sapiens (human)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria;
OC   Primates; Catarrhini; Hominidae; Homo.
XX
RN   [1]
RP   1-796
RA   Blum H., Bauersachs S., Mewes W., Gassenhuber J., Wiemann S.;
RT   ;
RL   Submitted (12-MAR-1999) to the EMBL/GenBank/DDBJ databases.
RL   MIPS, Am Klopferspitz 18a D-82152 Martinsried, GERMANY
```

```
XX
CC   Clone from S. Wiemann, sequenced by LMU within the cDNA
CC   sequencing consortium of the German Genome Project
CC   s1 sequence also available
CC   This clone is available at the RZPD in Berlin
CC   Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
CC   Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de
XX
FH   Key             Location/Qualifiers
FH
FT   source          1. .796
FT                   /db_xref="taxon:9606"
FT                   /organism="Homo sapiens"
FT                   /clone="DKFZp434A2223"
FT                   /clone_lib="434 (synonym: htes3). Vector pSport1; host
FT                   DH10B; sites NotI + SalI"
FT                   /dev_stage="adult"
FT                   /tissue_type="testis"
XX
SQ   Sequence 796 BP; 179 A; 210 C; 239 G; 158 T; 10 other;


alignment_scores:
            Quality:   74.50              Length:     66
              Ratio:  1.620                Gaps:      2
 Percent Similarity:  69.697   Percent Identity: 33.333

alignment_block:
 US-09-068-804-7 x HSM007904    ..

 Align seg 1/1  to: HSM007904  from: 1  to: 796

      42 GlyThrAspIleIleSerLeuSerGlnAlaAlaThrLysIleHisGlnAl 58
         |||   :::|||::::::|||             ::::::::::|||   ::
     176 GGGGAGGAGATCTTGGCTCTGCGT.........AAGCAGCTTCACAGCAC 216

      58 aGlnGlnThrLeuGlnSerThrProProIleSerGluGluAsnAsnAspG 75
         :||||||:::||||||   :::   ::::::::::||||||   :::|||
     217 CCAGCAGGCCCTGCAGTTTGCCAAGGCCATGGATGAGGAGCTGGAGGACC 266

      75 luArgThrLeuAlaArg......GlnGlnLeuThrSerSerLeuAsnAla 89
           :::|||||||||||||      ::::::   :::   ||||||   |||
     267 TGAAGACTCTGGCCAGGAGCCTGGAGGAACAGAATCGCAGCCTTCTGGCC 316

      90 LeuAlaLysSerGlyValSerLeuSerAlaGluGlnAsnGluAsnLeu 105
            |||:::   :::|||:::|||:::      ::::::   :::|||
     317 CAAGCCCGGCAGGCGGTGGGTCTGGCCCAGGGGAAGGAAGGTGCCCTC 364
```

This Page Blank (uspto)

OM of: US-09-068-804-14 to: EST:* out_format : pfs

Date: Oct 5, 1999 7:34 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09068804/runat_05101999_145134_10306/app_query.fasta.1
-DB=EST -QFMT=fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=pam150 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=escore -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfs -NORM=stat -USER=US09068804 -NCPU=6 -ICPU=3 -WAIT
-THREADS=1

Search information block:
Query: US-09-068-804-14
Query length: 513
Database: EST:*
Database sequences: 2546578
Database length: 986266752
Search time (sec): 940.790000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ..
gb_est11:AA238190	+	150.50	298.23	4.3e-08	620	! AA238190 mx11d09.r1 Soares mous
gb_est24:AI226302	+	150.50	294.61	6.8e-08	843	! AI226302 uf06e02.y1 Soares mous
gb_est25:AI304873	+	145.00	285.41	2.2e-07	681	! AI304873 qo19h01.x1 NCI_CGAP_Lu
gb_est20:AA870689	+	143.50	284.24	2.6e-07	574	! AA870689 vq24h06.r1 Barstead st
gb_est31:AI692576	+	143.00	280.90	3.9e-07	696	! AI692576 wd69e01.x1 NCI_CGAP_Lu
gb_est14:AA441056	+	142.00	280.29	4.3e-07	612	! AA441056 LD15830.5prime LD Dros
gb_est15:AA526445	+	132.50	262.03	4.4e-06	518	! AA526445 ni82a07.s1 NCI_CGAP_Pr
gb_est24:AI202762	+	131.50	259.09	6.5e-06	555	! AI202762 qi45d06.x1 NCI_CGAP_Br
gb_est22:AI048237	+	124.00	242.68	5.3e-05	576	! AI048237 ud71e08.y1 Sugano mous
gb_est30:AI664263	+	122.50	241.30	6.3e-05	494	! AI664263 ue83e05.r1 Soares mous
gb_est32:C82203	+	125.00	239.55	7.9e-05	900	! C82203 C82203 Leukocyte of Japa
gb_est8:AA035922	+	121.00	236.84	0.0001	550	! AA035922 mi69e09.r1 Soares mous
gb_est32:AI726050	+	121.50	236.12	0.0001	640	! AI726050 BNLGHi14043 Six-day Co
gb_est28:AI508841	+	119.50	235.08	0.0001	487	! AI508841 vc34g03.y1 Barstead MF
gb_est15:AA464338	+	121.00	234.69	0.0001	660	! AA464338 zx78f05.r1 Soares ovar
gb_est10:AA137348	+	120.50	233.97	0.0002	641	! AA137348 mg27d03.r1 Barstead MF
gb_est26:AU006346	+	121.50	233.83	0.0002	777	! AU006346 AU006346 Bombyx mori p
gb_est24:AI196724	+	120.00	232.51	0.0002	663	! AI196724 ui53h09.y1 Sugano mous
gb_est16:C50708	+	116.50	231.38	0.0002	388	! C50708 C50708 Yuji Kohara unpub
gb_est26:AI356774	+	119.50	231.25	0.0002	674	! AI356774 qy22h06.x1 NCI_CGAP_Br
gb_est19:AA764368	+	118.00	230.94	0.0002	528	! AA764368 vp08c03.r1 Soares mous
gb_est23:AI133923	-	119.00	230.70	0.0002	645	! AI133923 GH10981.5prime GH Dros
gb_est22:AI024728	-	119.50	230.40	0.0003	724	! AI024728 ov76e07.x1 Soares_test
em_est22:HSM008727	+	118.00	229.89	0.0003	577	! Al043877 Homo sapiens mRNA; EST
gb_est29:AI599536	+	117.50	229.74	0.0003	534	! AI599536 EST251239 Normalized r
gb_est20:AA817634	+	119.00	228.04	0.0003	808	! AA817634 LD24141.5prime LD Dros
gb_est25:AU039922	+	118.00	227.68	0.0004	696	! AU039922 AU039922 Dictyostelium
gb_est14:AA448018	-	116.00	227.39	0.0004	497	! AA448018 zw97a07.r1 Soares_tota
gb_est25:AI297917	+	116.00	226.77	0.0004	524	! AI297917 LP12362.5prime LP Dros
gb_est32:AI725967	+	117.00	226.54	0.0004	640	! AI725967 BNLGHi13737 Six-day Co
gb_est11:AA239491	+	114.50	223.15	0.0006	543	! AA239491 mw98h04.r1 Soares mous
gb_est28:AI512224	+	115.00	222.47	0.0007	630	! AI512224 LD43989.5prime LD Dros
gb_est29:AI552783	+	113.00	221.61	0.0008	472	! AI552783 mm24a03.y1 Stratagene
gb_est23:AI099196	+	113.00	220.16	0.0010	534	! AI099196 ue40c03.y1 Sugano mous
gb_est25:AI258873	+	114.00	220.08	0.0010	644	! AI258873 LP02171.5prime LP Dros
gb_est12:AA314039	+	112.50	220.04	0.0010	493	! AA314039 EST185892 Colon carcin
gb_est8:AA041943	+	113.50	219.37	0.0011	625	! AA041943 mj04e05.r1 Soares mous
gb_est14:AA392855	+	114.00	219.04	0.0011	703	! AA392855 LD11984.5prime LD Dros
gb_est8:AA000559	+	113.00	218.90	0.0011	594	! AA000559 mg23a06.r1 Soares mous
gb_est24:AI220156	+	112.50	218.20	0.0012	576	! AI220156 qg79b11.x1 Soares_NFL_
gb_est6:W13306	+	113.50	218.15	0.0012	693	! W13306 mb17c11.r1 Soares mouse
gb_est12:AA286039	-	110.50	217.31	0.0014	433	! AA286039 vc34g03.r1 Barstead MF
gb_est26:AU000589	+	112.50	216.09	0.0016	689	! AU000589 AU000589 Bombyx mor
gb_est26:AI332295	+	110.00	215.15	0.0018	475	! AI332295 fa97h10.y1 zebrafis
gb_est25:AU038955	+	110.00	212.98	0.0024	571	! AU038955 AU038955 Dictyostel

seq_name: gb_est11:AA238190

seq_documentation_block:
LOCUS AA238190 620 bp mRNA EST 03-MAR-1997
DEFINITION mx11d09.r1 Soares mouse NML Mus musculus cDNA clone IMAGE:679889 5' similar to gb:X63039 M.musculus RSP-1 mRNA for p33 protein (MOUSE);, mRNA sequence.
ACCESSION AA238190
NID g1862214
VERSION AA238190.1 GI:1862214
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 620)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1393259.

Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:419593
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 413.

FEATURES Location/Qualifiers
source 1. .620
/organism="Mus musculus"
/db_xref="taxon:10090"
/map="1"
/clone="IMAGE:679889"
/clone_lib="Soares mouse NML"
/tissue_type="Liver"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATCTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library constructed and normalized by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 164 a 181 c 137 g 137 t 1 others
ORIGIN

alignment_scores:
Quality: 150.50 Length: 213
Ratio: 1.275 Gaps: 10
Percent Similarity: 55.399 Percent Identity: 32.394

alignment_block:
US-09-068-804-14 x AA238190 ..

Align seg 1/1 to: AA238190 from: 1 to: 620

```
 9 AsnProValLeuAsnValGlyAlaSerGlyLeuThrThrLeuProAspAr 25
12 TCCAGTATGTTGGATGTC......AACGGCTTGTTCTCTCTA........ 47

25 gLeuProProHisIleThrThrLeuValIleProAspAsnAsnLeuThrS 42
48 .......GCCCACATCACACAACTGGTGCTCAGCCACAACAAGCTAACAA 90

42 erLeuProGluLeuProGluGlyLeuArgGluLeuGluValSerGly... 57
91 CTGTGCCACCAAATGTAGCGGAACTGAAGAACTTGGAGGTACTAAACTTC 140

58 ...AsnLeuGlnLeuThrSerLeuProSerLeuProGlnGlyLeuGlnLy 73
141 TTCAACAATCAGATCGAGGAACTGCCTACCCAGATCAGCAGCCTCCAGAA 190

73 sLeuTrpAlaTyrAsnAsnTrpLeuAlaSerLeuProThrLeuProProG 90
191 ACTCAAACACCTGAACCTTGGCATGAATAGGTTGAACACGCTGCCTCGAG 240

90 lyLeuGlyAspLeuAlaValSerAsnAsnGlnLeuThrSerLeuProGlu 106
241 GATTCGGCTCCCTCCCGGCT.....CTGGAGGTCCTGGACTTA 278

107 MetProProAlaLeuArgGluLeuArgValSerGlyAsn......AsnLe 121
279 ACTTACAACAACCTGAATGAACATTCTCTTCCGGGAAACTTCTTCTACCT 328

121 uThrSerLeuArgAla.....................LeuP 128
329 CACCACCCTGCGTGCACTCTATCTAAGCGACAACGATTTTGAAATCCTGC 378

128 roSerGly.....LeuGlnLysLeuTrpAlaTyrAsn 138
379 CTCCAGATATTGGGAAGCTCACAAAGTTGCAGATACTCAGCCTCAGGGAT 428

139 AsnArgLeuThrSerLeuPro.....GluMetSerProGlyLe 151
429 AATGACCTGATCTCACTGCCTAAGGAAATCGGGGAGCTGACCCAG...CT 475

151 uGlnGluLeuAspValSerHisAsnGlnLeuThrArgLeuProGlnSerL 168
476 GAAGGAGCTCCACATTCAGGGGAACCGCCTGNACGTTCTGCCTCCAGAGC 525

168 euThrGlyLeuSerSerAlaAlaArgValTyrLeuAspGlyAsnProLeu 184
526 TTGGCAACCTGGATCCT..... 542

185 SerValArgThrArgAspArgLeuCysGlyHisHisTrp 197
543 ACTGTCAGAAGCAAGTCTTCAAGCCGGGAACACCCCTGG 581
```

```
seq_name: gb_est24:AI226302

seq_documentation_block:
LOCUS       AI226302      843 bp    mRNA            EST       29-OCT-1998
DEFINITION  uf06e02.y1 Soares mouse mammary gland NMLMG Mus musculus cDNA clone
            IMAGE:1510586 5' similar to gb:X63039 M.musculus RSP-1 mRNA for p33
            protein (MOUSE);, mRNA sequence.
ACCESSION   AI226302
NID         g3809355
VERSION     AI226302.1  GI:3809355
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 843)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2045748.

            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:937438
            Seq primer: -40RP from Gibco
            High quality sequence stop: 423.
FEATURES             Location/Qualifiers
     source          1. .843
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /map="2"
                     /clone="IMAGE:1510586"
                     /clone_lib="Soares mouse mammary gland NMLMG"
                     /sex="female (lactating)"
                     /tissue_type="mammary gland"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; 1st strand cDNA was prepared from mammary
                     gland tissue from a lactating female, and was then primed
                     with a Not I - oligo(dT) primer. Double-stranded cDNA was
                     ligated to Eco RI adaptors (Pharmacia), digested with Not
                     I and cloned into the Not I and Eco RI sites of the
                     modified pT7T3 vector. Library is normalized. Library
                     was constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      233 a    239 c    170 g    175 t     26 others
ORIGIN


alignment_scores:
            Quality:  150.50              Length:    219
              Ratio:  1.234                 Gaps:     10
 Percent Similarity:  55.708   Percent Identity: 31.507

alignment_block:
 US-09-068-804-14 x AI226302   ..

 Align seg 1/1  to: AI226302  from: 1  to: 843

11 ValLeuAsnValGlyAlaSerGlyLeuThrThrLeuProAspArgLeuPr 27
 4 ATTCTGATGTTGGATGTCAACGGCTTGTTCTCTCTA.............. 39

27 oProHisIleThrThrLeuValIleProAspAsnAsnLeuThrSerLeuP 44
40 .GCCCACATCACACAACTGGTGCTCAGCCACAACAAGCTAACAACTGTGC 88

44 roGluLeuProGluGlyLeuArgGluLeuGluValSerGly......Asn 58
89 CACCAAATGTAGCGGAACTGAAGAACTTGGAGGTACTAAACTTCTTCAAC 138

59 LeuGlnLeuThrSerLeuProSerLeuProGlnGlyLeuGlnLysLeuTr 75
139 AATCAGATCGAGGAACTGCCTACCCAGATCAGCAGCCTCCAGAAACTCAA 188

75 pAlaTyrAsnAsnTrpLeuAlaSerLeuProThrLeuProProGlyLeuG 92
189 ACACCTGAACCTTGGCATGAATAGGTTGAACACGCTGCCTCGAGGATTCG 238

92 lyAspLeuAlaValSerAsnAsnGlnLeuThrSerLeuProGluMetPro 108
239 GCTCCCTCCCGGCT.....CTGGAGGTCCTGGACTTAACTTAC 276

109 ProAlaLeuArgGluLeuArgValSerGlyAsn......AsnLeuThrSe 123
```

```
          :::|||   |||      ::::::||||||         ||||||::
    277 AACAACCTGAATGAACATTCTCTTCCGGGAAACTTCTTCTACCTCACCAC 326

    123 rLeuArgAla.............................LeuProSerG 130
        :|||||||||                            ||||||::::
    327 CCTGCGTGCACTCTATCTAAGCGACAACGATTTTGAAATCCTGCCTCCAG 376

    130 ly..................LeuGlnLysLeuTrpAlaTyrAsnAsnArg 140
        ::                 ||||||   |||        :::|||
    377 ATATTGNGAAGCTCACAAAGTTGCAGATACTCAGCCTCAGGGATAATGAC 426

    141 LeuThrSerLeuPro............GluMetSerProGlyLeuGlnGl 153
        |||:::||||||||           |||::::::::::   |||:::||
    427 CTGATATCACTGCCTAAGGAAATCGGNGAGCTGACCCAG...CTGAAGGA 473

    153 uLeuAspValSerHisAsnGlnLeuThrArgLeuProGlnSerLeuThr. 169
        ||||::::::      |||:::||||||   ||||||:::   |||
    474 GCTCCACATTCAGGGGAACCGCCTGACCGTTCTGCCTCCAGAGCTTGGGC 523

    170 .....GlyLeuSerSerAlaAlaArgValTyr.................. 178
              |||::::::      :::|||:::
    524 ACTTGNATCTAACTGGTCAGAAGCAGGTCTTCANAGCAGAGAACAACCCC 573

    179 ..............LeuAspGlyAsnProLeuSerValArgThrArgAs 190
                      |||   :::::::::|||:::     |||   ::
    574 TGGTTACCCCGATTGCTGACCAGTTCCAGCTTGGCGTCTCCCACGTTNTC 623

    190 pArgLeu 192
        :::::::
    624 GAAAATA 630
```

seq_name: gb_est25:AI304873

seq_documentation_block:

```
LOCUS       AI304873      681 bp    mRNA            EST       01-FEB-1999
DEFINITION  qo19h01.x1 NCI_CGAP_Lu5 Homo sapiens cDNA clone IMAGE:1909009 3'
            similar to SW:RSU1_HUMAN Q15404 RAS SUPPRESSOR PROTEIN 1 ;, mRNA
            sequence.
ACCESSION   AI304873
NID         g3988562
VERSION     AI304873.1  GI:3988562
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 681)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Aug 21, 1998 this sequence version replaced.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 585   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 452.
FEATURES             Location/Qualifiers
     source          1. .681
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1909009"
                     /clone_lib="NCI_CGAP_Lu5"
                     /tissue_type="carcinoid"
                     /lab_host="DH10B"
                     /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; 1st strand cDNA was prepared from
                     neuroendocrine lung carcinoid, and was then primed with a
                     Not I - oligo(dT) primer. Double-stranded cDNA was ligated
                     to Eco RI adaptors (Pharmacia), digested with Not I and
                     cloned into the Not I and Eco RI sites of the modified
                     pT7T3 vector.  Library is normalized. Library was
                     constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      195 a    194 c    137 g    152 t      3 others
ORIGIN
```

alignment_scores:

```
            Quality:  145.00            Length:    221
              Ratio:  1.160               Gaps:     12
 Percent Similarity:  56.561   Percent Identity: 33.032
```

alignment_block:

US-09-068-804-14 x AI304873 ..

Align seg 1/1 to: AI304873 from: 1 to: 681

```
     17 SerGlyLeuThrThrLeuProAspArgLeuProProHisIleThrThrLe 33
        :::|||||||   ||||||              :::||||||||||   ||
      7 AACGGCCTCTTTACCTTA..............TCCCATATCACACAACT 41

     33 uValIleProAspAsnAsnLeuThrSerLeuProGluLeuProGluGlyL 50
        ||||:::::::::||||:::|||||||   :::|||          :::   |
     42 GGTCCTCAGCCATAACAAGCTAACAATGGTGCCACCGAACATCGCAGAAC 91

     50 euArgGluLeuGluValSerGly......AsnLeuGlnLeuThrSerLeu 64
        ||:::::::|||||||||   :::      |||   |||:::      |||
     92 TGAAGAATTTGGAGGTGCTCAACTTTTTTAATAACCAAATCGAGGAGCTG 141

     65 ProSerLeuProGlnGlyLeuGlnLysLeuTrpAlaTyrAsnAsnTrpLe 81
        |||:::      :::||||||||||||         |||      ::
    142 CCCACACAGATCAGTAGCCTTCAGAAACTCAAACACCTGAACCTTGGCAT 191

     81 uAlaSerLeuProThrLeuProProGlyLeuGlyAspLeuAlaValSerA 98
        ::::   |||   |||||||||   |||:::|||:::|||::::::
    192 GAACAGGCTGAACACTTTGCCACGAGGCTTCGGCTCCCTGCCAGCT.... 237

     98 snAsnGlnLeuThrSerLeuProGluMetProProAlaLeuArgGluLeu 114
              |||      |||              :::|||   |||
    238 ........CTTGAGGTTCTGGACTTGACTTACAACAACTTGAGCGAAAAT 279

    115 ArgValSerGlyAsn......AsnLeuThrSerLeuArgAla........ 126
        ::::::||||||        |||||||:::|||||||||
    280 TCTCTTCCTGGAAACTTCTTCTACCTGACCACCCTGCGTGCACTCTATCT 329

    127 .....................LeuProSerGly................. 130
                             |||||||:::::
    330 AAGTGACAACGATTNTGAAATCCTGCCGCCAGATATTGGGAAGCTCACAA 379

    131 ..LeuGlnLysLeuTrpAlaTyrAsnAsnArgLeuThrSerLeuPro... 145
          ||||||   |||        :::|||   |||:::|||||||||
    380 AGTTGCAGATACTCAGCCTTAGGGATAACGACCTGATCTCACTGCCTAAG 429

    146 .........GluMetSerProGlyLeuGlnGluLeuAspValSerHisAs 159
                ||||:::::::::   |||:::||||||::::::      ||
    430 GAAATCGGGGAGCTTACCCAG...CTTAAAGAGCTCCACATTCAGGGGAA 476

    159 nGlnLeuThrArgLeuProGlnSerLeuThr......GlyLeuSerSerA 174
        |:::|||||   ||||||:::   |||          :::||:::::
    477 CCGCCTCACCGTTCTGCCCCCAGAACTAGGAAACTTGGATTTAACTGGCC 526

    174 laAlaArgValTyr...LeuAspGlyAsnProLeuSerValArgThrArg 189
          ::::|||:::      ::::::||||||       :::   :::
    527 AGAAGCAGGTATTCAAAGCAGAGAACAATCCCTGGGTGACCCCCATTGCA 576
```

```
    190 AspArgLeu......Cys...GlyHisHisTrpProPheArgHisGlnAs 203
        |||::::::      |||   :::|||         :::   :::   ::
    577 GACCAGTTCCAGCNTTGCGTGTCCCATGTTTNIGAGTATATCCGTTCTGA 626

    203 pThrLeuArgTyr 207
        :|||   :::|||
    627 GACATACAAATAC 639

seq_name: gb_est20:AA870689

seq_documentation_block:
LOCUS       AA870689      574 bp    mRNA            EST       16-MAR-1998
DEFINITION  vq24h06.r1 Barstead stromal cell line MPLRB8 Mus musculus cDNA
            clone IMAGE:1095227 5' similar to gb:X63039 M.musculus RSP-1 mRNA
            for p33 protein (MOUSE);, mRNA sequence.
ACCESSION   AA870689
NID         g2966134
VERSION     AA870689.1  GI:2966134
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 574)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Jan 9, 1998 this sequence version replaced gi:936622.

            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:601459
            Trace considered overall poor quality
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .574
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1095227"
                     /clone_lib="Barstead stromal cell line MPLRB8"
                     /cell_line="C2C12 (undifferentiated)"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     [AATTCGGATCCTTG], digested with Not I and cloned into the
                     Not I and Eco RI sites of the modified pT7T3 vector.
                     Source undifferentiated tissue culture cell line C2C12.
                     Library constructed by Bob Barstead. The C2C12 cell line
                     (available from ATCC, catalog # CRL-1772) differentiates
                     rapidly, forming contractile myotubes and producing
                     characteristic muscle proteins.  "
BASE COUNT      154 a    156 c    140 g    124 t
ORIGIN


alignment_scores:
            Quality:  143.50          Length:    172
              Ratio:  1.449             Gaps:      7
```

```
 Percent Similarity:  57.558   Percent Identity: 34.302

alignment_block:
 US-09-068-804-14 x AA870689   ..

 Align seg 1/1  to: AA870689  from: 1  to: 574

     28 ProHisIleThrThrLeuValIleProAspAsnAsnLeuThrSerLeuPr 44
        :::|||||||||   ||||||:::::::::|||:::|||||:::::||
     57 GCCCACATCACACAACTGGTGCTCAGCCACAACAAGCTAACAGCTGTGCC 106

     44 oGluLeuProGluGlyLeuArgGluLeuGluValSerGly......AsnL 59
        |      :::   |||:::::|||||||||   :::      |||
    107 ACCAAATGTAGCGGAACTGAAGAACTTGGAGGTACTAAACTTCTTCAACA 156

     59 euGlnLeuThrSerLeuProSerLeuProGlnGlyLeuGlnLysLeuTrp 75
          |||:::      ||||||:::         :::|||||||||||
    157 ATCAGATCGAGAAACTGCCTACCCAGATCAGCAGCCTCCAGAAACTCAAA 206

     76 AlaTyrAsnAsnTrpLeuAlaSerLeuProThrLeuProProGlyLeuGl 92
            |||      ::::::   |||   |||||||||   |||:::||
    207 CACCTGAACCTTGGCATGGATAGGTTGAACACGCTGCCTCGAGGATTCGG 256

     92 yAspLeuAlaValSerAsnAsnGlnLeuThrSerLeuProGluMetProP 109
        |:::|||||:::           |||      |||       :
    257 CTCCCTCGCGGCT...........CTGGAGGTCCTGGACTTAACTTACA 294

    109 roAlaLeuArgGluLeuArgValSerGlyAsn......AsnLeuThrSer 123
        :::::|||           ::::::||||||         |||||||||
    295 GCAACCTGAATGGACATTCTCTTCCGGGAAACTTCTTCTACCTCACCAGC 344

    124 LeuArg.............................AlaLeuProSerGl 130
        :::|||                                |||||:::::
    345 ATGCGTTGACTCTATCTAAGCGACGACGATTTTGAAATCCTGCCTCCAGA 394

    130 yLeuGlnLysLeu......TrpAlaTyr...........AsnAsnArgL 141
        ::::   ||||||      |||                  :::|||   |
    395 TATTGGGAAGCTCCAAAGTTGGCAGATACTCAGCCTCAGGGATAATGAGC 444

    141 euThrSerLeuProGluMetSerProGly.........LeuGlnGluLeu 154
        ||   |||||||||           |||          |||:::||||||
    445 TGATGTCACTGCCTAAGGAAATCGGGGGAGTGACCCAGCTGGAAGAGCTC 494

    155 AspValSerHisAsnGlnLeuThrArgLeuProGlnSerLeuThrGlyLe 171
        :::|||      :::|||:::|||   |||::::::   |||   :::||
    495 CACGTTCAGGGGGACCAAATGACCGTTCTGGCTCCAGAGCTTGGAGACTT 544

    171 uSerSerAlaAlaArg 176
        |:::::::::::::::
    545 GGGACCAACTGGTCAC 560

seq_name: gb_est31:AI692576

seq_documentation_block:
LOCUS       AI692576      696 bp    mRNA            EST       02-JUN-1999
DEFINITION  wd69e01.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:2336856 3'
            similar to TR:O75473 O75473 ORPHAN G PROTEIN-COUPLED RECEPTOR HG38.
            ;, mRNA sequence.
ACCESSION   AI692576
NID         g4969916
VERSION     AI692576.1  GI:4969916
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 696)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3187661.
```

```
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco
            High quality sequence stop: 469.
FEATURES              Location/Qualifiers
     source           1. .696
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /map="17q21"
                      /clone="IMAGE:2336856"
                      /clone_lib="NCI_CGAP_Lu24"
                      /tissue_type="carcinoid"
                      /lab_host="DH10B"
                      /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                      modified polylinker; Plasmid DNA from the normalized
                      library NCI_CGAP_Lu5 was prepared, and ss circles were
                      made in vitro. Following HAP purification, this DNA was
                      used as tracer in a subtractive hybridization reaction.
                      The driver was PCR-amplified cDNAs from a pool of 5,000
                      clones made from the same library (cloneIDs
                      1414920-1417991 and 1520904-1522439). Subtraction by Bento
                      Soares and M. Fatima Bonaldo. "
BASE COUNT      157 a    198 c    172 g    166 t      3 others
ORIGIN


alignment_scores:
           Quality:  143.00              Length:    206
             Ratio:  1.144                 Gaps:      9
 Percent Similarity:  60.680   Percent Identity: 27.670

alignment_block:
 US-09-068-804-14 x AI692576   ..

 Align seg 1/1  to: AI692576  from: 1  to: 696

       4 CysLeuAsnAsnGlyAsnProValLeuAsnValGlyAlaSerGlyLeuTh 20
         |||      :::|||:::      ::::::   ::::::   ||||||||
      82 TGCAGCTGCGACGGCGACCGTCGGGTGGACTGCTCCGGGAAGGGGCTGAC 131

      20 rThrLeuProAspArgLeuProProHisIleThrThrLeuValIleProA 37
         |:::::|||:::   |||::::::   :::   :::|||   |||:::
     132 GGCCGTGCCCGAGGGGCTCAGCGCCTTCACCCAAGCGCTGGATATCAGTA 181

      37 spAsnAsnLeuThrSerLeuProGluLeuProGluGlyLeuArgGluLeu 53
         ||||||:::|||   |||||||||      ::::::::::::::::::
     182 TGAACAACATTACTCAGTTGCCAGAA......GATGCATTTAAGAACTTT 225

      54 GluValSerGlyAsnLeuGlnLeuThrSer......LeuProSerLeu.. 67
                 :::||||||||||:::::      |||:::   :::
     226 CCTTTTCTAGAAGAGCTACAATTGGCGGGCAACGACCTTTCTTTTATCCA 275

      68 .ProGlnGlyLeuGlnLysLeuTrpAlaTyrAsnAsnTrpLeuAlaSerL 84
          |||::::::|||      |||
     276 CCCAAAGGCCTTGTCTGGGTTG.......................... 297

      84 euProThrLeuProProGlyLeuGlyAspLeuAlaValSerAsnAsnGln 100
                       |||      |||::::::   |||||||||
     298 ............AAAGAACTCAAAGTTCTAACGCTCCAGAATAATCAG 333

     101 LeuThrSerLeuProGluMetPro...........ProAlaLeuArgGl 113
         |||:::::::::|||      :::          :::|||||||:::
     334 TTGAAAACAGTACCCAGTGAAGCCATTCGAGGGCTGAGTGCTTTGCAGTC 383

     113 uLeuArgValSerGlyAsnAsnLeuThrSerLeuArgAlaLeuPro.... 128
         ||||||:::::::::|||:::::|||||||:::   :::   :::
     384 TTTGCGTTTAGATGCCAACCATATTACCTCAGTCCCCGAGGACAGTTTTG 433

     129 .......SerGlyLeuGlnLysLeuTrpAlaTyrAsnAsnArgLeuThr 142
                 |||:::   ||||||      :::|||   ||||||
     434 AAGGACTTGTTCAGTTACGGCATCTGTGGCTGGATGACAACAGCTTGACG 483

     143 SerLeuProGluMetSer...........ProGlyLeuGlnGluLeuAs 155
         :::|||      :::           |||   ||||||:::|||
     484 GAGGTGCCTGTGCACCCNCTCAGCAATCTGCCCACCCTACAGGCGCTGAC 533

     155 pValSerHisAsnGlnLeuThrArgLeuProGln...SerLeuThrGlyL 171
         ::::::   |||:::::::::   :::|||:::   :::::::|||:::|
     534 CCTGGCTCTCAACAAGATCTCAAGCATCCCTGACTTTGCATTTACCAACC 583

     171 euSerSerAlaAlaArgValTyrLeuAspGlyAsnProLeuSerValArg 187
         ||||||||   :::   ::::::|||::::::|||          :::|||
     584 TTTCAAGCCTGGTAGTTCTGCATCTTCATAACAAT......AAAATTAGA 627

     188 ThrArgAspArgLeuCys 193
         :::   ::::::   |||
     628 AGCCTGAGTCAACACTGT 645

seq_name: gb_est14:AA441056

seq_documentation_block:
LOCUS       AA441056      612 bp    mRNA            EST       28-NOV-1998
DEFINITION  LD15830.5prime LD Drosophila melanogaster embryo BlueScript
            Drosophila melanogaster cDNA clone LD15830 5prime, mRNA sequence.
ACCESSION   AA441056
NID         g2152934
VERSION     AA441056.1  GI:2152934
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 612)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1394853.

            Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 158  row: C  column: 6
            High quality sequence stop: 526.
FEATURES              Location/Qualifiers
     source           1. .612
                      /organism="Drosophila melanogaster"
                      /db_xref="BDGP_EST:BDcln014969"
                      /db_xref="taxon:7227"
                      /clone="LD15830"
                      /clone_lib="LD Drosophila melanogaster embryo BlueScript"
                      /sex="male and female"
                      /dev_stage="0 to 24 hours mixed stage embryonic"
                      /lab_host="SOLR"
                      /note="Organ: embryo; Vector: BlueScript SK; Site_1:
                      EcoRI; Site_2: XhoI; Constructed using Stratagene ZAP-cDNA
                      Synthesis kit. Oligo dT-primed and directionally cloned at
                      EcoRI and XhoI in BlueScript SK(+/-)"
BASE COUNT      140 a    181 c    156 g    135 t
ORIGIN
```

```
alignment_scores:
            Quality:  142.00              Length:    160
              Ratio:  1.464                 Gaps:      8
 Percent Similarity:  60.625   Percent Identity: 33.125

alignment_block:
 US-09-068-804-14 x AA441056   ..

 Align seg 1/1  to: AA441056  from: 1  to: 612

      25 ArgLeuProProHisIle.........ThrThrLeuValIleProAspAs 38
         ||||||||||||:::|||          :::||||||      |||
     171 CGCCTGCCGCCGGATATACAAAACTTTGAAAATCTCGTTGAGCTCGACGT 220

      38 nAsnLeuThrSerLeuProGluLeuProGluGlyLeuArgGluLeuGluV 55
          :::   :::::::::|||:::::::|||:::::::::::::::::|||:::
     221 TTCACGCAATGATATACCCGACATTCCCGACGATATCAAGCATCTGCAG. 269

      55 alSerGlyAsnLeuGlnLeuThr..............SerLeuProSer 66
                :::|||||||::::::               :::::::|||
     270 ........AGCCTGCAGGTGGCTGATTTTAGTTCGAACCCGATACCCAAG 311

      67 LeuProGlnGlyLeuGlnLysLeu..............TrpAlaTyrAs 78
         ||||||   |||:::   :::|||                ::::   ||
     312 TTGCCTTCGGGCTTCTCCCAGCTAAAGAACCTGACCGTCCTGGGACTCAA 361

      78 nAsnTrpLeuAlaSerLeuProThrLeuProProGlyLeuGlyAspLeuA 95
         |:::         ||||||:::|||||||||:::::::::|||:::|||:
     362 CGAC......ATGTCGCTCACCACATTGCCAGCGGATTTTGGCAGCCTGA 405

      95 laValSerAsnAsnGlnLeuThrSerLeuProGluMetProProAlaLeu 111
         ::           ||||||   ||||||         :::       |||
     406 CT............CAATTGGAATCCCTGGAGCTTCGCACGAATCTGCTC 443

     112 ArgGluLeuArgValSerGlyAsnAsnLeuThrSerLeuArgAlaLeuPr 128
         :::::::|||      :::   ::::::||||||   |||:::   |||
     444 AAGCATCTGCCCGAGACAATAAGCCAGCTGACGAAGCTGAAGCGTCTGGA 493

     128 oSerGly.........LeuGlnLysLeuTrpAlaTyrAsnAsnArgLeuT 142
             |||         ::::::   |||   :::|||   :::   |||
     494 TCTGGGCGACAACGAGATCGAGGATCTGCCACCCTATTTGGGTTATCTG. 542

     142 hrSerLeuProGluMetSerProGlyLeuGlnGluLeuAspValSerHis 158
                              |||||||||:::||||||   :::::|||
     543 ....................CCGGGTCTGCACGAGCTTTGGTTGGATCAC 572

     159 AsnGlnLeuThrArgLeuProGlnSerLeu 168
         |||||||||   |||||||||:::   |||
     573 AATCAGCTGCAGCGATTGCCGCCGGAGCTG 602

 seq_name: gb_est15:AA526445

 seq_documentation_block:
 LOCUS       AA526445      518 bp    mRNA            EST       20-AUG-1997
 DEFINITION  ni82a07.s1 NCI_CGAP_Pr20 Homo sapiens cDNA clone IMAGE:983316, mRNA
             sequence.
 ACCESSION   AA526445
 NID         g2268514
 VERSION     AA526445.1  GI:2268514
 KEYWORDS    EST.
 SOURCE      human.
   ORGANISM  Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
             Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE   1  (bases 1 to 518)
   AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
   TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
             Tumor Gene Index
   JOURNAL   Unpublished (1997)
 COMMENT     On Jan 24, 1995 this sequence version replaced gi:634327.

             Contact: Robert Strausberg, Ph.D.
             Tel: (301) 496-1550
             Email: Robert_Strausberg@nih.gov
             Tissue Procurement: Paul H. Duray, M.D., Rodrigo F. Chuaqui, M.D.,
             Michael R. Emmert-Buck, M.D., Ph.D.
              cDNA Library Preparation: David B. Krizman, Ph.D.
              cDNA Library Arrayed by: Greg Lennon, Ph.D.
              DNA Sequencing by: Washington University Genome Sequencing Center
              Clone distribution: NCI-CGAP clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             www-bio.llnl.gov/bbrp/image/image.html

             Seq primer: -40m13 fwd. ET from Amersham
             High quality sequence stop: 272.
 FEATURES             Location/Qualifiers
      source          1. .518
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:983316"
                      /clone_lib="NCI_CGAP_Pr20"
                      /sex="male"
                      /tissue_type="prostate metastasis, liver"
                      /lab_host="DH10B"
                      /note="Vector: pAMP10; mRNA made from metastatic prostate
                      lesion of the liver, cDNA made by  oligo-dT priming.
                      Non-directionally cloned.  Size-selected on agarose  gel,
                      average insert size 600 bp.  Library made by D. Krizman,
                      NIH."
 BASE COUNT      135 a    170 c     98 g    115 t
 ORIGIN


 alignment_scores:
             Quality:  132.50              Length:    136
               Ratio:  1.559                 Gaps:      1
  Percent Similarity:  62.500   Percent Identity: 34.559

 alignment_block:
  US-09-068-804-14 x AA526445   ..

  Align seg 1/1  to: AA526445  from: 1  to: 518

       30 IleThrThrLeuValIleProAspAsnAsnLeuThrSerLeuProGluLe 46
          :::|||:::|||          |||::::::|||||||||||||:::
       11 CTGACTAGTTTACAGTCACACGATGGTGACCTGACTAGTTTACAATCACA 60

       46 uProGluGlyLeuArgGluLeuGluValSerGlyAsnLeuGlnLeuThrS 63
                ::::|||       |||:::   |||   :::   :::|||||||
       61 CGACCCTGACCTGACTAGTTTACAA...TCACACGACCCTGACCTGACTA 107

       63 erLeuProSerLeuProGlnGlyLeuGlnLysLeuTrpAlaTyrAsnAsn 79
          |||||:::|||      :::::::|||      |||   :::::::::
      108 GTTTACAATCACACGACCCTGACCTGACTAGTTTACAATCACACGATCCT 157

       80 TrpLeuAlaSerLeuProThrLeuProProGlyLeuGlyAspLeuAlaVa 96
             |||:::||||||::::::      |||:::|||   :::|||
      158 GACCTGACTAGTTTACAATCACACGACCCTGACCTGACTAGTTTACAATC 207

       96 lSerAsnAsnGlnLeuThrSerLeuProGluMetProProAlaLeuArgG 113
             :::   :::|||||||||||||:::         |||:::|||
      208 ACACGACCCTGACCTGACTAGTTTACAGTCACACGATCCTGACCTGACTA 257

      113 luLeuArgValSerGlyAsnAsnLeuThrSerLeuArgAlaLeuProSer 129
          |||:::      :::   :::||||||||||||:::::      :::
      258 GTTTACAACCACACGACCCTGACCTGACTAGTTTACAATCACACGACCCT 307

      130 GlyLeuGlnLysLeuTrpAlaTyrAsnAsnArgLeuThrSerLeuProGl 146
          :::|||      |||   ::::::         ||||||||||||:::
      308 GACCTGACTAGTTTACAATCACACTATCCTGACCTGACTAGTTTACAATC 357

      146 uMetSerProGlyLeuGlnGluLeuAspValSerHisAsnGlnLeuThrA 163
          :::|||:::|||      |||:::          :::||||||
      358 ACACGACCCTGACCTGACTAGTCTACAATCCCACGGACCTGACCTGACTA 407

      163 rgLeuPro 165
```

```
          ||||||
     408 GTTIACCG 415

seq_name: gb_est24:AI202762

seq_documentation_block:
LOCUS       AI202762      555 bp    mRNA            EST       01-DEC-1998
DEFINITION  qi45d06.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:1859435 3'
            similar to TR:P70193 P70193 INTEGRAL MEMBRANE GLYCOPROTEIN. ;, mRNA
            sequence.
ACCESSION   AI202762
NID         g3755368
VERSION     AI202762.1  GI:3755368
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 555)
  AUTHORS   NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute / National Institute of Neurological
            Disorders and Stroke, Brain Tumor Genome Anatomy Project
            (CGAP/BTGAP), Tumor Gene Index
  JOURNAL   Unpublished (1998)
COMMENT     On Jan 14, 1998 this sequence version replaced gi:1797843.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
            Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 628   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 453.
FEATURES             Location/Qualifiers
     source          1. .555
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1859435"
                     /clone_lib="NCI_CGAP_Brn25"
                     /tissue_type="anaplastic oligodendroglioma"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATAGGTTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     Library is normalized, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT      128 a    154 c    151 g    122 t
ORIGIN


alignment_scores:
            Quality:  131.50            Length:    189
              Ratio:  1.134               Gaps:     11
 Percent Similarity:  61.376   Percent Identity: 29.630

alignment_block:
 US-09-068-804-14 x AI202762   ..

 Align seg 1/1  to: AI202762  from: 1  to: 555
```

```
      12 LeuAsnValGlyAlaSerGlyLeuThrThrLeuProAspArgLeuProPr 28
         |||:::   |||:::   ||||||::::::||||||:::   ||||||::
      13 CTGGACTGCGGTGGGCGCGGGCTGGCTGCGTTGCCCGGGGACCTGCCCTC 62

      28 oHisIleThrThrLeuValIleProAspAsnAsnLeu..............  40
         :   :::   :::|||   ::::::   |||:::|||
      63 CTGGACGCGGAGCCTAAACCTGAGTTACAACAAACTCTCTGAGATTGACC 112

      41 ..ThrSerLeuProGluLeuProGluGlyLeuArgGluLeuGluValSer 56
           :::::::::   :::||||||:::   |||:::|||:::   ::::::
     113 CTGCTGGTTTTGAGGACTTGCCGAAC...CTACAGGAAGTGTACCTCAAT 159

      57 GlyAsnLeuGlnLeuThrSerLeuProSerLeuProGlnGly........  70
         :::|||   :::||||||::::::|||||||||      :::
     160 AATAAT...GAGTTGACAGCGGTACCATCCCTGGGCGCTGCTTCATCACA 206

      71 ................LeuGln.............................  72
                         ||||||
     207 TGTCGTCTCTCTCTTTCTGCAGCACAACAAGATTCGCAGCGTGGAGGGGA 256

      73 ..LysLeuTrpAlaTyrAsnAsnTrpLeuAlaSerLeuProThrLeuPro 88
           :::|||   ||||||               ||||||   :::|||
     257 GCCAGCTGAAGGCCTAC............CTTTCCTTAGAAGTGTTAGAT 294

      89 ProGlyLeuGlyAspLeuAlaValSerAsnAsnGlnLeuThrSerLeuPr 105
           :::|||::::::      :::::::::   :::   |||:::::::||
     295 CTGAGTTTGAACAAC......ATCACGGAAGTGCGGAACACCTGCTTTCC 338

     105 oGluMetProProAlaLeuArgGluLeuArgValSerGlyAsn....... 119
         |:::   ||||||   ::::::|||||   ::::::||||||
     339 ACACGGACCGCCT...ATAAAGGAGCTCAACCTGGCAGGCAATCGGATTG 385

     120 ........AsnLeuThrSerLeuArgAlaLeuProSerGlyLeuGlnLys 133
                 :::|||   ::::::   :::|||:::   :::|||   :::
     386 GCACCCTGGAGTTGGGAGCATTTGATGGTCTGTCACGGTCGCTGCTAACT 435

     134 LeuTrpAlaTyrAsnAsnArgLeuThrSerLeuProGluMetSer..... 148
         |||:::      :::||||||:::|||   ||||||      :::
     436 CTTCGCCTGAGCAAAAACAGGATCACCCAGCTTCCTGTAAGAGCATTCAA 485

     149 ....ProGlyLeuGlnGluLeuAspValSerHisAsnGlnLeuThrArgL 164
             |||   |||   :::||||||::::::::::|||::::::   ||||
     486 GCTACCCAGGCTGACACAACTGGACCTCAATCGGAACAGGATT...CGGC 532

     164 euProGlnSerLeuThr 169
         ||   ::::::||||||
     533 TGATAGAGGGCCTCACC 549

seq_name: gb_est22:AI048237

seq_documentation_block:
LOCUS       AI048237      576 bp    mRNA            EST       08-JUL-1998
DEFINITION  ud71e08.y1 Sugano mouse liver mlia Mus musculus cDNA clone
            IMAGE:1451366 5' similar to SW:A2GL_HUMAN P02750 LEUCINE-RICH
            ALPHA-2-GLYCOPROTEIN ;, mRNA sequence.
ACCESSION   AI048237
NID         g3296524
VERSION     AI048237.1  GI:3296524
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 576)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On May 5, 1995 this sequence version replaced gi:797624.
```

```
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:924682
            Seq primer: custom primer used
            High quality sequence stop: 469.
FEATURES              Location/Qualifiers
     source           1. .576
                      /organism="Mus musculus"
                      /strain="C57BL"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:1451366"
                      /clone_lib="Sugano mouse liver mlia"
                      /sex="female"
                      /dev_stage="adult"
                      /lab_host="DH10B"
                      /note="Organ: liver; Vector: pME18S-FL3; Site_1: DraIII
                      (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
                      was primed with an oligo(dT) primer
                      [ATGTGGCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was
                      ligated to a  DraIII adaptor [TGTTGGCCTACTGG], digested
                      and cloned into distinct DraIII sites of the pME18S-FL3
                      vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should
                      be used to isolate the cDNA insert.  Size selection was
                      performed to exclude fragments <1.5kb.  Library
                      constructed by Dr. Sumio Sugano (University of Tokyo
                      Institute of Medical Science).  Custom primers  for
                      sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
                      primer CGACCTGCAGCTCGAGCACA."
BASE COUNT       87 a    210 c    169 g    109 t      1 others
ORIGIN


alignment_scores:
            Quality:  124.00            Length:    174
              Ratio:  1.265               Gaps:      7
 Percent Similarity:  56.322   Percent Identity: 29.885

alignment_block:
 US-09-068-804-14 x AI048237   ..

 Align seg 1/1  to: AI048237  from: 1  to: 576

       1 MetArgAspCysLeuAsn...AsnGlyAsnProValLeuAsnValGlyAl 16
         :::::::::|||||||      :::::::::          :::|||:::
     114 CTCAAGGAATGCCTGATACTGCAGTCGGCTGAGGGTAGCACCGTCTCCTG 163

      16 aSerGlyLeuThrThrLeuProAspArgLeuProProHisIleThrThrL 33
            |||   |||   :::|||:::   |||||||:::::::::::::   |
     164 CCATGGTCCCACCGAGTTTCCGAGCTCCCTCCCTGCCGACACTGTCCATC 213

      33 euValIleProAspAsnAsnLeuThrSerLeuProGluLeuPro..... 47
         ||   :::      :::||||||||||   ||||||:::   :::
     214 TGTCGGTGGAATTCTCCAACCTGACGCAGCTGCCCGCCGCCGCCCTGCAA 263

      48 ......GluGlyLeuArgGluLeuGluValSerGlyAsnLeuGlnLeuTh 62
               ||||||||||||||||:::::::|||:::|||   :::|||
     264 GGGTGTCCGGGTCTGCGGGAGCTGCACCTCTCGAGCAAT...CGCCTGCA 310

      62 rSerLeu...ProSerLeuProGlnGlyLeuGlnLysLeuTrpAlaTyrA 78
          :::|||   |||   |||         :::::::::|||:::|||   :
     311 GGCGCTGTCTCCCGAGTTGCTGGCGCCCGTGCCCCGGCTGCGCGCCCTGG 360

      78 snAsnTrpLeuAlaSerLeuProThrLeuProProGlyLeuGlyAspLeu 94
         ::         :::::|||   :::||||||||||||         |||
     361 ATCTGACCCGCAATGCTCTCCGCAGCCTGCCCCCCGGG.........TTG 401

      95 AlaValSerAsnAsnGlnLeuThrSerLeuProGluMetProProAlaLe 111
            :::::::::::|||:::::::|||                   ||
     402 TTCAGCACCTCGGCCAACCTGAGCACCCTGGTCCTGAGGGAGAACCAGCT 451

     111 uArgGluLeuArgValSerGlyAsnAsnLeuThrSerLeuArgAlaLeuP 128
         |||||||      ||||||::::::   |||   :::|||   |||
     452 GCGGGAG......GTGAGCGCGCAATGGCTTCAGGGCCTGGACGCC.... 491

     128 roSerGlyLeuGlnLysLeuTrpAlaTyrAsnAsnArgLeuThrSerLeu 144

     491 ..................................................... 491

     145 ProGluMetSerProGlyLeuGlnGluLeuAspValSerHisAsnGlnLe 161
                           |||   :::|||||||:::::::::|||||||
     492 ..................CTGNGCCACTTGGACCTGGCGGAGAACCAGCT 523

     161 uThrArgLeuProGlnSerLeu 168
         |:::   ||||||   :::|||
     524 GAGTTCGCTGCCCTCCGGGCTC 545

seq_name: gb_est30:AI664263

seq_documentation_block:
LOCUS       AI664263      494 bp    mRNA            EST       10-MAY-1999
DEFINITION  ue83e05.r1 Soares mouse uterus NMPu Mus musculus cDNA clone
            IMAGE:1497728 5' similar to gb:X63039 M.musculus RSP-1 mRNA for p33
            protein (MOUSE);, mRNA sequence.
ACCESSION   AI664263
NID         g4767846
VERSION     AI664263.1  GI:4767846
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 494)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
            Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
            Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     On May 18, 1998 this sequence version replaced gi:3138145.

            Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:935332
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 366.
FEATURES              Location/Qualifiers
     source           1. .494
                      /organism="Mus musculus"
                      /db_xref="taxon:10090"
                      /map="1p"
                      /clone="IMAGE:1497728"
                      /clone_lib="Soares mouse uterus NMPu"
                      /sex="female"
                      /dev_stage="adult"
                      /lab_host="DH10B"
                      /note="Organ: uterus; Vector: pT7T3D-Pac (Pharmacia) with
                      a modified polylinker; 1st strand cDNA was prepared from
                      pregnant mouse uterus, and was then primed with a Not I -
                      oligo(dT) primer. Double-stranded cDNA was ligated to Eco
                      RI adaptors (Pharmacia), digested with Not I and cloned
                      into the Not I and Eco RI sites of the modified pT7T3
                      vector.  Library is normalized. Library was constructed by
```

```
                    Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      137 a    141 c    112 g    104 t
ORIGIN


alignment_scores:
            Quality:  122.50              Length:    162
              Ratio:  1.303                 Gaps:      7
 Percent Similarity:  58.025   Percent Identity: 33.951

alignment_block:
 US-09-068-804-14 x AI664263   ..

 Align seg 1/1  to: AI664263  from: 1  to: 494

      34 ValIleProAspAsnAsnLeuThrSerLeuProGluLeuProGluGlyLe 50
         |||:::::::::|||:::|||||||:::::|||          :::   ||
       3 GTGCTCAGCCACAACAAGCTAACAACTGTGCCACCAAATGTAGCGGAACT 52

      50 uArgGluLeuGluValSerGly......AsnLeuGlnLeuThrSerLeuP 65
         |:::::|||||||||   ::       :::   |||:::      ||||
      53 GAAGAACTTGGAGGTACTAAACTTCTTCGACAATCAGATCGAGGAACTGC 102

      65 roSerLeuProGlnGlyLeuGlnLysLeuTrpAlaTyrAsnAsnTrpLeu 81
         ||:::         :::|||||||||||||        |||      :::
     103 CTACCCAGATCAGCAGCCTCCAGAAACTCAAACACCTGAACCTTGGCATG 152

      82 AlaSerLeuProThrLeuProProGlyLeuGlyAspLeuAlaValSerAs 98
         :::   |||   |||||||||   |||:::|||:::|||:::::
     153 AATAGGTTGAACACGCTGCCTCGAGGATTCGGCTCCCTCCCGGCT..... 197

      98 nAsnGlnLeuThrSerLeuProGluMetProProAlaLeuArgGluLeuA 115
            |||      |||:::            :::|||   |||
     198 .......CTGGAGGTCCTGGCCTTAACTTACAACAACCTGGATGAACATT 240

     115 rgValSerGlyAsn......AsnLeuThrSerLeuArgAla......... 126
          ::::::|||||         ||||||:::|||||||||
     241 CTCTTCCGGGAAACTTCTTCTACCTCACCACCCTGCGTGCACTCTATCTA 290

     127 .....................LeuProSerGly................. 130
                              ||||||::::::
     291 AGCGACAACGATTTTGAAATCCTGCCTCCAGATATAGGGAAGCTCACGAA 340

     131 .LeuGlnLysLeuTrpAlaTyrAsnAsnArgLeuThrSerLeuPro.... 145
          ||||||   |||         :::|||   |||:::|||||||||
     341 GTTGCAGATACTCAGCCTCAGGGATAATGACCTGATCTCACTGCCTAAGG 390

     146 ........GluMetSerProGlyLeuGlnGluLeuAspValSerHisAsn 159
               |||:::::::::   |||:::||||||::::::      :::
     391 AAATCGGGGAGCTGACCCAG...CTGAAGGAGCTCCACATTCAGGGGAAG 437

     160 GlnLeuThrArgLeuProGlnSerLeuThrGlyLeu 171
         :::||||||   ||||||:::   |||   :::|||
     438 CGCCTGACCGTACTGCCTCCAGAGCTTGAGAACTTG 473

seq_name: gb_est32:C82203

seq_documentation_block:
LOCUS       C82203        900 bp    mRNA            EST       10-JUN-1999
DEFINITION  C82203 Leukocyte of Japanese flounder i Paralichthys olivaceus cDNA
            clone WF7-4, mRNA sequence.
ACCESSION   C82203
NID         g5039439
VERSION     C82203.1  GI:5039439
KEYWORDS    EST.
SOURCE      flounder.
  ORGANISM  Paralichthys olivaceus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
            Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Percomorpha;
            Pleuronectiformes; Pleuronectoidei; Bothidae; Paralichthys.
REFERENCE   1  (bases 1 to 900)
  AUTHORS   Nam,B., Katagiri,T., Hong,Y., Hirono,I. and Aoki,T.
  TITLE     A survey of expressed genes in the leukocyte of Japanese flounder
            (Paralychthys olivaceus) infected with Hirame rhabdo virus
  JOURNAL   Unpublished (1999)
COMMENT     On Mar 10, 1998 this sequence version replaced gi:2948001.

            Contact: Takashi Aoki
            Genetics and Biochemistry
            Tokyo University of Fisheries
            Konan, 4-5-7, Minato 108, Japan
            Tel: 03-5463-0556
            Fax: 03-5463-0690
            Email: ad95217@s4201.tokyo-u-fish.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .900
                     /organism="Paralichthys olivaceus"
                     /db_xref="taxon:8255"
                     /clone="WF7-4"
                     /clone_lib="Leukocyte of Japanese flounder i"
                     /sex="Female"
                     /cell_type="leukocyte infected with Hirame r"
BASE COUNT      182 a    289 c    243 g    186 t
ORIGIN


alignment_scores:
            Quality:  125.00              Length:    155
              Ratio:  1.344                 Gaps:      8
 Percent Similarity:  60.000   Percent Identity: 32.258

alignment_block:
 US-09-068-804-14 x C82203   ..

 Align seg 1/1  to: C82203  from: 1  to: 900

      43 LeuProGluLeuProGluGlyLeuArgGluLeuGluValSer........ 56
         |||   |||:::|||   :::   |||:::   ::::::|||
     299 CTGAAGGAGTTTCCCCGCAGCGCCCGCAACTACGACCTGTCCGACATCAC 348

      57 ....GlyAsnLeuGlnLeuThrSerLeuProSerLeuProGln....... 69
             :::::::|||      :::   |||      ||||||:::
     349 ACACGCAGACCTGTCTAAGAACCGTCTCTGTGAGTTGCCGGAGGAGCTCT 398

      70 ...........GlyLeuGlnLysLeuTrpAlaTyrAsnAsnTrpLeuAla 82
                   :::|||::::::|||      |||:::|||
     399 GTCAGTTCATCTCTCTGGAGACACTGAGCCTTTACCACAAT......... 439

      83 SerLeuProThrLeuProProGlyLeuGlyAsp.................. 93
         ::::::   :::|||:::::::::||||||:::
     440 GGGATGCGTTCACTGTCCTCCAGCCTGGGCAACCTCCAGGCTCTGACCTA 489

      94 .LeuAlaValSerAsnAsnGlnLeuThrSerLeuProGlu.........M 107
          |||:::::::|||   |||   |||:::|||||||||
     490 CCTCAACCTCAGTCGTAATCTCCTGTCCAGTTTACCCCCGTCGGTGTTTC 539

     107 etProProAlaLeuArgGluLeuArgValSerGlyAsnAsnLeuThrSer 123
             |||   ||||||   |||   ||||||::::|||:::|||:::|||
     540 AGCTTCCTCTCCTGCGAGTGCTCATCGTCAGCAACAACAAGCTGTCTTCA 589

     124 LeuArgAlaLeuProSerGlyLeuGlnLysLeuTrpAlaTyrAsn..... 138
         |||   |||         :::|||     |||:::      :::
     590 CTGCCCGCCTCCATATACTCCCTCACACACCTCCGACAGCTGGACGTGAG 639

     139 ....AsnArgLeuThrSerLeuPro...GluMetSerPro......GlyL 151
             |||   |||   :::|||||||   |||:::::::::          |
     640 CTGCAATGAGCTGCAGTGTTTGCCTGCAGAGCTCGGTCAGTTGGAGTGTC 689

     151 euGlnGluLeuAspValSerHisAsnGlnLeuThrArgLeuProGlnSer 167
         ||::::::|||::::::   :::|||||||||||||   ||||||::::::
     690 TGAGGGACTTAAACCTGAGGAGGAACCAGCTCACCACACTGCCTGAAGAT 739

     168 LeuThrGlyLeuSer 172
         ::::::   |||:::
     740 ATATCCGAGCTCCCT 754
```

```
seq_name: gb_est8:AA035922

seq_documentation_block:
LOCUS       AA035922      550 bp    mRNA            EST       26-AUG-1996
DEFINITION  mi69e09.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
            clone IMAGE:468808 5' similar to gb:X63039 M.musculus RSP-1 mRNA
            for p33 protein (MOUSE);, mRNA sequence.
ACCESSION   AA035922
NID         g1509065
VERSION     AA035922.1  GI:1509065
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 550)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Apr 14, 1993 this sequence version replaced gi:716426.

            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:282624
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 475.
FEATURES             Location/Qualifiers
     source          1. .550
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:468808"
                     /clone_lib="Soares mouse embryo NbME13.5 14.5"
                     /sex="unknown"
                     /tissue_type="embryo"
                     /dev_stage="13.5-14.5dpc total fetus"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGGAAATTTTTTTTTTTTTTTTTTTTTT
                     T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
                     14.5dpc embryos [total RNA provided by  Minoru Ko, Wayne
                     State Univ., from 2 ]; double-stranded cDNA was ligated to
                     Eco RI adaptors (Pharmacia), digested with Not I and
                     cloned into the Not I and Eco RI sites of the modified
                     pT7T3 vector. Library went through one round of
                     normalization, and was constructed by Bento Soares and
                     M.Fatima Bonaldo.  "
BASE COUNT      147 a    155 c    124 g    124 t
ORIGIN


alignment_scores:
            Quality:  121.00            Length:    194
              Ratio:  1.120               Gaps:      6
 Percent Similarity:  55.670   Percent Identity: 25.258

alignment_block:
 US-09-068-804-14 x AA035922   ..

 Align seg 1/1  to: AA035922  from: 1  to: 550
```

```
 22 LeuProAspArgLeuProProHisIleThrThrLeuValIleProAspAs 38
    :::|||   ::::::|||   |||:::      :::::::::::::::||
 36 TTTCCCTTGCAGGTCCCTAATCACACCCACCAGGTGTTGCTCGCTGCGAA 85

 38 nAsnLeuThrSerLeuProGluLeuProGluGlyLeuArgGluLeuGlu. 54
    |     :::||||||      |||   |||       ||||||   :::
 86 CATGTCCAAGTCACTGAAGAAGCTGGTGGAGGAGAGCAGGGAGAAGAACC 135

 55 ............................ValSerGlyAsnLeuGlnLeu 61
                                 :::|||:::   |||:::::
136 AGCCGGAAGTGGACATGAGTGACAGGGGTATCTCCAGTATGTTGGATGTC 185

 62 ThrSerLeuProSerLeuProGlnGlyLeuGlnLysLeuTrpAlaTyrAs 78
    :::::|||   ||||||::::::      |||   :::   ::::::||
186 AACGGCTTGTTCTCTCTAGCCCACATCACACAACTGGTGCTCAGCCACAA 235

 78 nAsnTrpLeuAlaSerLeuProThrLeuProProGlyLeuGlyAsp.... 93
    |:::          |||:::|||:::||||||:::::::::::::
236 CAAG.............CTAACAACTGTGCCACCAAATGTAGCGGAACTGA 273

 94 ..............LeuAlaValSerAsnAsnGlnLeuThrSerLeuPro 105
                   |||:::      |||||||||:::      ||||||
274 AGAACTTGGAGGTACTAAACTTCTTCAACAATCAGATCGAGGAACTGCCT 323

106 GluMetProProAlaLeuArgGluLeuArgValSerGlyAsnAsnLeuTh 122
             :::::::|||:::   |||:::      :::   ::::::::
324 ACCCAGATCAGCAGCCTCCAGAAACTCAAACACCTGAACCTTGGCATGAA 373

122 rSerLeuArgAlaLeuProSerGlyLeuGlnLysLeuTrpAlaTyrAsnA 139
    :   |||   :::||||||   |||:::      |||
374 TAGGTTGAACACGCTGCCTCGAGGATTCGGCTCCCTC.............. 410

139 snArgLeuThrSerLeuProGluMetSerProGlyLeuGlnGluLeuAsp 155
                                 |||:::|||:::   ||||||
411 ............................CCGGCTCTGGAGGTCCTGGAC 431

156 ValSerHisAsnGlnLeuThr......ArgLeuProGlnSerLeuThrGl 170
    :::::::::|||:::|||:::         ||||||   ::::::
432 TTAACTTACAACAACCTGAATGAACATTCTCTTCCGGGAAACTTCTTCTA 481

170 yLeuSerSerAlaAlaArgValTyrLeuAspGlyAsnProLeuSerValA 187
     |||::::::       |||||||||:::::::::   |||
482 CCTCACCACCCTGCGTGCAGTCTATCTAAGCGATACGATTTTG....... 524

187 rgThrArgAspArgLeuCysGlyHisHisTrp 197
              :::   |||:::::::::|||
525 ...........AAATCATGCTCCAGATATTGG 545
```

```
seq_name: gb_est32:AI726050

seq_documentation_block:
LOCUS       AI726050      640 bp    mRNA            EST       11-JUN-1999
DEFINITION  BNLGHi14043 Six-day Cotton fiber Gossypium hirsutum cDNA 5' similar
            to (AC002329) predicted leucine-rich protein [Arabidopsis
            thaliana], mRNA sequence.
ACCESSION   AI726050
NID         g5044902
VERSION     AI726050.1  GI:5044902
KEYWORDS    EST.
SOURCE      upland cotton.
  ORGANISM  Gossypium hirsutum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
REFERENCE   1  (bases 1 to 640)
  AUTHORS   Blewitt,M., Matz,E.C., Davy,D.F. and Burr,B.
  TITLE     ESTs from developing cotton fiber
  JOURNAL   Unpublished (1999)
COMMENT     On May 18, 1998 this sequence version replaced gi:3137655.

            Contact: Ben Burr
            Biology Department
```

Brookhaven National Laboratory
Upton, NY 11973, USA
Tel: 516-344-3396
Fax: 516-344-3407
Email: burr@bnlux1.bnl.gov
Seq primer: T3 Primer.

```
FEATURES             Location/Qualifiers
     source          1. .640
                     /organism="Gossypium hirsutum"
                     /cultivar="Acala Maxxa"
                     /db_xref="taxon:3635"
                     /map="4"
                     /clone_lib="Six-day Cotton fiber"
                     /tissue_type="immature fiber"
                     /dev_stage="Six days post anthesis"
                     /lab_host="XL1-Blue"
                     /note="Vector: pBluescript II KS+"
BASE COUNT      184 a    124 c    155 g    177 t
ORIGIN


alignment_scores:
            Quality:  121.50             Length:    155
              Ratio:  1.413                Gaps:      6
 Percent Similarity:  55.484   Percent Identity: 31.613

alignment_block:
 US-09-068-804-14 x AI726050   ..

 Align seg 1/1  to: AI726050  from: 1  to: 640

      38 AsnAsnLeuThrSerLeuProGluLeuProGluGlyLeuArgGluLeuGl 54
         |||   |||:::::::|||||||||            :::      |||||
      12 AACCGGCTCAAAGCCCTCCCAGAAGCAGTTGGGAAGATAGAAACCTTAGA 61

      54 uValSerGlyAsnLeuGlnLeuThrSerLeuProSerLeuProGlnGlyL 71
         ||||      :::::::::   :::::::::      ||||||      :
      62 GGTTTTA...TCCGTGCGCTATAACAACATCAAACAATTACCTACGACGA 108

      71 euGlnLysLeuTrpAlaTyrAsnAsnTrpLeuAlaSerLeuProThrLeu 87
         ::      |||
     109 TGTCATCTTTG..................................... 119

      88 ProProGlyLeuGlyAspLeuAlaValSerAsnAsnGlnLeuThrSerLe 104
            ::::::|||   :::|||:::||||||   |||:::|||   |||||
     120 ...GCAAACCTGAAGGAATTAGATGTCAGTTTCAATGAGCTTGAGTCTTT 166

     104 uProGlu.........MetProProAlaLeuArgGluLeuArgValSerG 118
         |||||||              ::::::::::|||      :::   |||   |
     167 GCCTGAGAGTTTATGTTTTGCCACCACACTAGTCAAGATGAACGTA...G 213

     118 lyAsnAsnLeuThrSerLeuArgAlaLeuProSerGlyLeuGlnLysLeu 134
         ||||||||:::::::::::|||:::||||||   ::::::   :::|||
     214 GCAACAATTTTGCTGATATGCGATCCCTACCAAGGTCTATTGGAAATCTT 263

     135 TrpAlaTyrAsnAsnArgLeuThrSerLeuProGluMetSerProGlyLe 151
                                        ||||||         ||
     264 ..............................GAGATG.........CT 271

     151 uGlnGluLeuAspValSerHisAsnGlnLeuThrArgLeuProGlnSerL 168
         |:::|||||||||:::|||:::||||||:::      ||||||:::|||:
     272 TGAAGAGCTGGACATTAGCAATAATCAGATCCGGGTCCTCCCAGACTCTT 321

     168 euThrGlyLeuSerSerAlaAlaArgValTyrLeuAspGlyAsnProLeu 184
         ::      |||:::            :::   :::|||   |||||||||
     322 TTAGGATGTTAACACGACTTCAAGTTCTACGCGTGGATCAAAATCCTCTG 371

     185 SerValArgThrArg 189
            |||   :::|||
     372 GAAGTGCCTCCTAGG 386

seq_name: gb_est28:AI508841
```

```
seq_documentation_block:
LOCUS        AI508841      487 bp    mRNA            EST       12-MAR-1999
DEFINITION   vc34g03.y1 Barstead MPLRB1 Mus musculus cDNA clone IMAGE:776500 5'
             similar to WP:AC7.2 CE07652 ;, mRNA sequence.
ACCESSION    AI508841
NID          g4407746
VERSION      AI508841.1  GI:4407746
KEYWORDS     EST.
SOURCE       house mouse.
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
             Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 487)
  AUTHORS    Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
             Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
             Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
             Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
             Waterston,R. and Wilson,R.
  TITLE      The WashU-NCI Mouse EST Project 1999
  JOURNAL    Unpublished (1999)
COMMENT      On Jun 5, 1998 this sequence version replaced gi:3187528.

             Contact: Marra M/WashU-NCI Mouse EST Project 1999
             Washington University School of Medicine
             4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
             Tel: 314 286 1800
             Fax: 314 286 1810
             Email: mouseest@watson.wustl.edu
             This clone is available royalty-free through LLNL ; contact the
             IMAGE Consortium (info@image.llnl.gov) for further information.
             MGI:469356
             This read is a RESEQUENCE of a previously sequenced mouse clone
             This read has been verified (found to hit its original self in the
             correct orientation)
             Possible reversed clone: similarity on wrong strand
             Seq primer: -40RP from Gibco
             High quality sequence stop: 407.
FEATURES             Location/Qualifiers
     source          1. .487
                     /organism="Mus musculus"
                     /strain="BALB/c"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:776500"
                     /clone_lib="Barstead MPLRB1"
                     /sex="mixed"
                     /tissue_type="Kidney"
                     /dev_stage="6 weeks"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     [CATGGATTCGGTACC], digested with Not I and cloned into the
                     Not I and Eco RI sites of the modified pT7T3 vector.
                     Library constructed by Bob Barstead."
BASE COUNT      140 a    120 c    101 g    126 t
ORIGIN


alignment_scores:
            Quality:  119.50             Length:    147
              Ratio:  1.374                Gaps:      7
 Percent Similarity:  59.184   Percent Identity: 31.293

alignment_block:
 US-09-068-804-14 x AI508841   ..

 Align seg 1/1  to: AI508841  from: 1  to: 487

      50 LeuArgGluLeuGluValSerGlyAsnLeuGlnLeuThrSerLeuProSe 66
         |||      |||::::::   :::|||   |||||||||||||||||
     107 TTAAGTAAGCTGAATATGAAGGACAAT...CAGTTAACATCACTTCCT.. 151
```

```
     66 rLeuProGlnGlyLeuGlnLysLeuTrpAlaTyrAsnAsnTrpLeuAlaS 83
                                 :::::::::::::|||
    152 .........................TTGGATTTTGGAACTTGG....... 169

     83 erLeuProThrLeuProProGlyLeuGlyAspLeuAlaValSerAsnAsn 99
                        :::::::::   :::|||:::::::::::::|||
    170 ..................ACCAGTATGGTAGAATTGAATTTAGCGACTAAT 202

    100 GlnLeuThrSerLeuProGluMetProProAlaLeuArgGluLeuArgVa 116
        |||||||||   :::|||||||      ::::::|||      |||   ||
    203 CAGCTCACAAAGATCCCAGAGGATGTGTCTGGTCTCGTTTCCCTTGAGGT 252

    116 l......SerGlySerAsnLeuThrSerLeuArgAlaLeuProSerGlyL 131
        |         :::||||||||||       |||:::   ||||||   ||||
    253 TCTGATCTTATCGAACAACCTT......CTAAAGAAGCTGCCCCACGGCC 296

    131 euGlnLysLeuTrpAlaTyr.................AsnAsnArgLeu 141
        ||   :::|||:::                      :::|||:::|||
    297 TTGGGAACCTCAGAAAGCTACGAGAGCTGGACCTGGAGGAGAACAAGCTG 346

    142 ThrSerLeuProGluMetSerPro.........GlyLeuGlnGluLeuAs 155
           ||||||||||:::      :::         :::||||||   |||
    347 GAGTCCTTGCCCAATGAGATCGCGTATCTCAAGGATCTGCAGAAATTAGT 396

    155 pValSerHisAsnGlnLeuThrArgLeuProGlnSerLeuThrGlyLeuS 172
         :::::::::|||||||||:::   ||||||:::::::::      |||:
    397 CTTGACAAACAACCAGTTGAGCACGCTTCCCAGAGGCATCGGTCACCTTA 446

    172 erSerAlaAlaArgValTyrLeuAspGlyAsnProLeuSer 185
        :::::   :::::::::   |||:::   |||   |||:::
    447 CCAACCTCACGCACCTTGGTCTTGGAGAGAACCTGCTCACT 487
```

seq_name: gb_est15:AA464338

seq_documentation_block:

```
LOCUS       AA464338      660 bp    mRNA            EST       10-JUN-1997
DEFINITION  zx78f05.r1 Soares ovary tumor NbHOT Homo sapiens cDNA clone
            IMAGE:809889 5' similar to gb:M14219 BONE PROTEOGLYCAN II PRECURSOR
            (HUMAN);, mRNA sequence.
ACCESSION   AA464338
NID         g2189222
VERSION     AA464338.1  GI:2189222
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 660)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J.,
            Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-Merck EST Project 1997
  JOURNAL   Unpublished (1997)
COMMENT     On Jun 27, 1996 this sequence version replaced gi:1184736.

            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 425.
FEATURES             Location/Qualifiers
     source          1. .660
                     /organism="Homo sapiens"
                     /db_xref="GDB:6039734"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:809889"
                     /clone_lib="Soares ovary tumor NbHOT"
                     /sex="Female"
                     /tissue_type="ovarian tumor"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: ovary; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGGTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library constructed by Bento Soares and
                     M.Fatima Bonaldo. "
BASE COUNT      202 a    156 c    141 g    161 t
ORIGIN
```

alignment_scores:

```
            Quality:  121.00              Length:    201
              Ratio:  1.071                 Gaps:      9
 Percent Similarity:  56.219   Percent Identity: 25.373
```

alignment_block:

US-09-068-804-14 x AA464338 ..

Align seg 1/1 to: AA464338 from: 1 to: 660

```
     26 LeuProProHisIleThrThrLeuValIleProAspAsnAsnLeuThrSe 42
        ||||||||||:::::|||   |||   :::::::::|||:::::|||
      1 CTTCCCCCTGACACAACTCTGCTAGACCTGCAAAACAACAAAATAACCGA 50

     42 rLeuProGluLeuProGluGlyLeuArgGluLeuGluValSerGlyAsnL 59
         :::::::::          ::::::::::::|||   :::      :::|
     51 AATCCAAGAT......GGAGACTTTAAGAACCTGAAGACGCTTCACGCAT 94

     59 euGlnLeuThrSerLeuProSerLeuProGlnGlyLeuGlnLysLeuTrp 75
        ||   |||::::::
     95 TGATTCTTGTCAAC................................... 108

     76 AlaTyrAsnAsnTrpLeuAlaSerLeuProThrLeuProProGly..... 90
                           :::   ::::::::::::::::||||||
    109 ..................AATAAAATTAGCAAAGTTAGTCCTGGAGCATT 140

     91 ................LeuGlyAspLeuAlaValSerAsnAsnGlnLeuT 102
                        |||      |||   :::|||:::|||||||||:
    141 TACACCTTTGGTAGAGTTGGAACGACTTTATCTGTCCAAGAATCAGCTGA 190

    102 hrSerLeuProGlu...MetProProAlaLeuArgGluLeuArgValSer 117
        ::   |||||||||   ||||||   :::|||:::|||||||||:::
    191 AGGAATTGCCAGAAAAAATGCCCAAAACTCTTCAGGAGCTGCGTGCCCAT 240

    118 GlyAsnAsnLeuThrSerLeuArgAlaLeuPro.................. 128
            |||::::::|||   :::|||   ::::::
    241 GAGAATGAGATCACCAAAGTGCGAAAAGTTACTTTCAATGGACTGAACCA 290

    128 ..................................................... 128

    291 GATGATTGTCATAGAACTGGGCACCAATCCGCTGAAGAGCTCAGGAATTG 340

    129 ..............SerGlyLeuGlnLysLeuTrpAlaTyr......... 137
                        |||::::::||||||   :::|||
    341 AAAATGGGGCTTTCCAGGGAATGAAGAAGCTC...TCCTACATCCGCATT 387

    138 ...AsnAsnArgLeuThrSerLeuProGlu...MetSerProGlyLeuGl 152
        ::::::   :::|||||||:::|||:::   ::::::|||:::|||
    388 GCTGATACCAATATCACCAGCATTCCTCAAGGTCTTCCTCCTTCCCTTAC 437

    152 nGluLeuAspValSerHisAsnGlnLeuThrArgLeu...ProGlnSerL 168
         ||||||:::::::::   |||:::::::::|||:::   :::   ||||
    438 GGAATTACATCTTGATGGCAACAAAATCAGCAGAGTTGATGCAGCTAGCC 487

    168 euThrGlyLeuSerSerAlaAlaArgValTyrLeuAspGlyAsnProLeu 184
        ||:::||||||::::::   |||::::::   |||:::   |||::::::
```

```
488 TGAAAGGACTGAATAATTTGGCTAAGTTGGGATTGAGTGTCAACAGCATC 537

185 Ser 185
    |||
538 TCT 540
```

This Page Blank (uspto)

OM of: US-09-068-804-7 to: Issued_Patents_NA:* out_format : pfs

Date: Oct 5, 1999 7:11 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=x1p
-Q=/cgn2_1/USPTO_spool/US09068804/runat_05101999_145135_10326/app_query.fasta.1
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=pam150
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=escore
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09068804
-NCPU=6 -ICPU=3 -WAIT -THREADS=1

Search information block:
Query: US-09-068-804-7
Query length: 344
Database: Issued_Patents_NA:*
Database sequences: 192659
Database length: 52021692
Search time (sec): 101.650000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ..
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-317-401E-1 +			79.50	143.15	1.39	2877 !
/cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:PCT-US92-06965A-18 +			72.50	137.13	3.00	11
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-609-230A-8 +			77.50	135.45	3.73	3751 !
/cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:PCT-US95-14418-3 +			73.50	134.63	4.14	1794
/cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:PCT-US95-15327-3 +			73.50	134.63	4.14	1794
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-320-559-1 +			82.50	131.73	6.01	14255 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-327-392-1 +			82.50	131.73	6.01	14255 !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-306-691B-55 +			82.50	131.73	6.01	14255
/cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:PCT-US94-04496-1 +			82.50	131.73	6.01	1425
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-977-434-9 +			73.50	130.80	6.77	2505 !
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-458-819-9 +			73.50	130.80	6.77	2505 !
/cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:PCT-US91-07035-9 +			73.50	130.80	6.77	2505
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-648-657-14 +			71.50	130.70	6.85	1686 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-384-490-30 +			73.50	130.20	7.31	2640 !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-459-383-30 +			73.50	130.20	7.31	2640 !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-463-081B-34 +			72.50	129.71	7.78	2249 !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-461-379A-34 +			72.50	129.71	7.78	2249 !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-462-390B-34 +			72.50	129.71	7.78	2249 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-073-384C-3 +			73.00	129.64	7.85	2504 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-254-359A-3 +			73.00	129.64	7.85	2504 !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-483-043-3 +			73.00	129.64	7.85	2504 !
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-481-238-3 +			73.00	129.64	7.85	2504 !
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-471-066B-3 +			73.00	129.64	7.85	2504 !
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-484-956-3 +			73.00	129.64	7.85	2504 !
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-757-653-3 +			73.00	129.64	7.85	2504 !
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-599-491-3 +			73.00	129.64	7.85	2504 !
/cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:PCT-US95-14418-1 +			73.50	128.54	9.04	3048
/cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:PCT-US95-15327-1 +			73.50	128.54	9.04	3048
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-977-434-7 +			72.50	128.48	9.12	2505 !
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-458-819-7 +			72.50	128.48	9.12	2505 !
/cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:PCT-US91-07035-7 +			72.50	128.48	9.12	2505
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-276-452A-24 +			70.50	128.35	9.27	1690 !
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-798-744-24 +			70.50	128.35	9.27	1690 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-080-255-5 +			68.50	126.54	11.70	1321 !
/cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:PCT-US93-05857-5 +			68.50	126.54	11.70	1321
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-533-306A-3 +			72.00	125.68	13.05	2887 !
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-742-923A-3 +			72.00	125.68	13.05	2887 !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-761-258-1 +			78.50	125.67	13.08	10763 !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-977-306-1 +			78.50	125.67	13.08	10763 !
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-533-306A-5 +			71.50	125.38	13.57	2680 !
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-742-923A-5 +			71.50	125.38	13.57	2680 !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-463-081B-30 +			69.00	121.38	22.65	2289 !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-461-379A-30 +			69.00	121.38	22.65	2289
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-462-390B-30 +			69.00	121.38	22.65	2289
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-466-390-1 +			69.00	121.03	23.70	2360

seq_name: /cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-317-401E-1

seq_documentation_block:
; Sequence 1, Application US/08317401E
; Patent No. 5922561
; GENERAL INFORMATION:
; APPLICANT: Thompson, Sheryl Ann
; APPLICANT: Yaver, Debbie Sue
; TITLE OF INVENTION: GENES ENCODING SIGNAL RECOGNITION PARTICLE OF
; TITLE OF INVENTION: ASPERGILLUS NIGER
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5922561o No. 5922561disk of No. 5922561th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/317,401E
; FILING DATE: 03-October-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Harrington, James J.
; REGISTRATION NUMBER: 38,711
; REFERENCE/DOCKET NUMBER: 4248.000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2877 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Aspergillus niger
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join (126..203, 253..1776)
US-08-317-401E-1

alignment_scores:
Quality: 79.50 Length: 346
Ratio: 0.437 Gaps: 15
Percent Similarity: 52.601 Percent Identity: 19.075

alignment_block:
US-09-068-804-7 x US-08-317-401E-1 ..

Align seg 1/1 to: US-08-317-401E-1 from: 1 to: 2877

```
      3 AsnIleGlnAsnTyrSerAlaSerProHis......ProGlyIleValAl 17
        :::|||:::   |||::::::  ::::::      |||:::|||||||
    643 AAGATCCCCTACTACGGTAGCCTGACGCAAACCGACCCCGCCATCGTGGC 692

     17 aGluArgProGlnThrProSerAlaSerGluHisValGluThrAlaValV 34
        |:::        :::           |||:::  |||:::   :::|
    693 AGCCGAGGGTGTGGCCAAGTTCAAGAAGGAGCGTTTCGAAATCATCATTG 742

     34 alProSerThrThrGluHisArgGlyThrAsp......IleIleSerLeu 48
        ||   ::::::     |||:::     :::      ::::::   :::
    743 TCGATACCAGTGGTCGTCATAAGCAGGAAGAAGAGCTCTTCACCGAAATG 792
```

```
 49 SerGlnAlaAlaThrLysIleHisGlnAlaGlnGlnThrLeuGlnSerTh 65
    :::|||      |||   :::   ::::::|||   :::|||
793 ACCCAGATTCAGACCGCCGTCACCCCCGACCAGACCATCCTCGTCCTCGA 842

 65 rProProIleSerGluGluAsnAsnAspGluArgThrLeu.AlaArgGln 81
     ::::::|||:::::::::::::::::::::::   :::    :::|||:::
843 CAGCACCATCGGTCAGGCTGCCGAAGCCCAGTCCTCCGCCTTCAAGGCCA 892

 82 GlnLeuThrSerSerLeuAsnAlaLeuAlaLysSerGlyValSerLeuSe 98
    :::   ||||||
893 CCGCAGACTTCG..................................... 904

 98 rAlaGluGlnAsnGluAsnLeuArgSerAlaPheSerAlaProThrSerA 115
      |||::::::   :::   |||          ::::::::::::   :
905 ....GAGCCATCATCATCACCAAGACGGATGGTCACGCCGCAGGTGGTGG 950

115 laLeuPheSerAlaSerPro............................. 121
    ::||||||::::::||||||
951 TGCTATTTCCGCCGTCGCCGCCACACACACTCCCATTATCTACCTCGGTA 1000

122 ..........................MetAlaGlnProArgThrThrIl 129
                              ::::::   ||||||::::::
1001 CCGGTGAGCACTTGATGGACCTGGAACGTTTCGAGCCCAAGGCCTTCATC 1050

129 eSerAspAlaGluIleTrp..AspMetValSer..............Gl 140
      ::::::  :::|||  ||||||::::::               ||
1051 CAGAAGCTCCTGGGTATGGGTGATATGGCCGGACTGGTAGAGCACGTACA 1100

140 nAsnIleSerAlaIleGlyAspSerTyrLeuGlyValTyrGluAsnValV 157
    |::::::::::      ::::::|||          :::|||   ::::::
1101 AGCCGTGACCAAGGACTCGGCCTCCGCCAAGGAAACATACAAGCACATCT 1150

157 al...AlaValTyrThr......AspPheTyrGlnAlaPheSerAspIle 170
      ::::::::::||      ||||||   ::::::::::::::|||
1151 CAGAAGGTATCTACACGCTGCGCGACTTCCGCGAGAACATCACCTCCATC 1200

171 ..............LeuSerLysMetGlyGlyTrpLeuLeuProGlyLy 182
                 ||||||||::::::|||   ::::::||||||
1201 ATGAAGATGGGTCCTCTCTCCAAGCTCTCCGGC...ATGATTCCCGGTCT 1247

182 sAspGlyAsnThrValLysLeuAspValThrSerLeuLysAsnAspLeuA 199
     ::::::   |||:::   ||||||      :::             |||
1248 CTCCAACCTGACCGCGGGACTTGATGACGAAGACGGCTCCATGAAGCTCC 1297

199 snSerLeuValAsnLysTyrAsn..............GlnIleAsnSer 210
      ::::::      ::::::               ::::::::::::
1298 GCCGCATGATCTACATTTTCGACAGCATGACGGCCGCCGAACTCGACGGC 1347

211 AsnThrValLeuProAlaGlnSerGlySerGlyValLysValAla.. 226
    :::      :::|||   :::|||:::::::      |||:::::::|||
1348 GACGGCAAGATGTTCGTCGAACAGCCTAGCCGCATGGTCCGGATCGCTTG 1397

227 ............ThrGluAlaGluAlaArgGlnTrpLeuSerGluLeuA 239
                |||      |||:::   :::   ||||||:::
1398 CGGAAGCGGTACCACCGTCCGCGAAGTCGAAGACCTGCTCTCCCAGCACC 1447

239 snLeuProAsnSerCysLeuLysSerTyrGlySer.............. 250
      :::   ::::::      |||      |||:::
1448 GCATGATGGCTGGTATGGCCAAGCGCGTCGGCGGACAGAAGAAGCAGATG 1497

251 ......................GlyTyrValValThrValAspLeuTh 259
                           |||                 :::|||::
1498 CAGCGTGCCCAGAACATGCTCAAGGGCGGTAACAAGGAACAGCAGCTCGC 1547

259 rProLeuGlnLysMetValGlnAspIleAspGlyLeuGlyAla..... 273
    ::::::::||||||   :::   ::::::::::|||   |||:::
1548 CGCCATGCAGAAGCGCATGGCCGCCATGGGTGGTGCTGGCGGTGGTGGAT 1597

274 ..ProGlyLysAspSerLysLeuGluMetAspAsn 284
      ||||||:::   ::::::   :::|||::::::
1598 TCCCCGGTATGCCCGGCATGGGCGACATGGCCAAG 1632
```

```
seq_name: /cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:PCT-US92-06965A-18

seq_documentation_block:
; Sequence 18, Application PC/TUS9206965A
;  GENERAL INFORMATION:
;    APPLICANT:  DEVARE, S.
;    APPLICANT:  DESAI, S.
;    APPLICANT:  DAILEY, S.
;    TITLE OF INVENTION:  HCV SYNTHETIC PEPTIDE FROM NS1 REGION
;    NUMBER OF SEQUENCES:  35
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ABBOTT LABORATORIES
;      STREET:  ONE ABBOTT PARK ROAD
;      CITY:  ABBOTT PARK
;      STATE:  ILLINOIS
;      COUNTRY:  U.S.
;      ZIP:  60065-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/06965A
;      FILING DATE:  19920821
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  POREMBSKI, PRISCILLA E.
;      REGISTRATION NUMBER:  33,207
;      REFERENCE/DOCKET NUMBER:  4834PC.02
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  708-937-6365
;      TELEFAX:  708-937-9556
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1179 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  circular
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1179
PCT-US92-06965A-18


alignment_scores:
           Quality:   72.50              Length:    283
             Ratio:  0.462                Gaps:     18
 Percent Similarity:  55.477   Percent Identity: 19.081

alignment_block:
 US-09-068-804-7 x PCT-US92-06965A-18   ..

 Align seg 1/1  to: PCT-US92-06965A-18  from: 1  to: 1179

  17 AlaGluArgProGlnThrProSerAlaSerGluHisValGluThrAlaVa 33
     |||||||||      ::::::::::::   |||:::|||:::   |||||
 124 GCCGAGCGCATCATCGTGGCAACCGATCATGAGGATGTTGCCCGCGCCGT 173

  33 lValProSerThrThrGlu.............................. 39
     |   ::::::      |||
 174 TGAAGCCGCTGGCGGTGAAGTATGTATGACGCGCGCCGATCATCAGTCAG 223

  40 .....HisArgGlyThrAspIleIle.........SerLeuSerGlnAla 51
          :::|||   ::::::::::::         ::::::|||::::::
 224 GAACAGAACGTCTGGCGGAAGTTGTCGAAAAATGCGCATTCAGCGACGAC 273

  52 AlaThrLysIleHis...GlnAlaGlnGlnThrLeuGlnSerThrProPr 67
     ::::::   ::::::   |||:::::::::::::::   ::::::::::
 274 ACGGTGATCGTTAATGTGCAGGGTGATGAACCGATGATCCCTGCGACAAT 323

  67 oIleSerGluGluAsnAsnAspGluArgThrLeuAlaArgGlnGlnLeuT 84
```

```
     324 CATTCGTCAGGTTGCTGATAAC.........CTCGCTCAGCGTCAGGTG. 363

      84 hrSerSerLeuAsnAlaLeuAlaLysSerGlyValSerLeuSerAlaGlu 100
     364 .............GGTATGGCGACTCTGGCGGTGCCAATCCACAATGCG 399

     101 GlnAsn............GluAsnLeuArgSerAlaPheSerAlaProTh 113
     400 GAAGAAGCGTTTAACCCGAATGCGGTGAAAGTGGTTCTCGACGCTGAAGG 449

     113 rSerAlaLeuPheSerAlaSerProMetAlaGlnProArgThrThrIleS 130
     450 GTATGCACTGTACTTC..................TCTCGCGCCACCATTC 481

     130 erAspAlaGluIleTrp......AspMetValSerGlnAsnIleSerAla 144
     482 CT............TGGGATCGTGATCGTTTTGCAGAAGGCCTTGAAACC 519

     145 IleGlyAspSerTyr.........LeuGlyValTyrGluAsnValValAl 158
     520 GTTGGCGATAACTTCCTGCGTCATCTTGGTATTTATGGC........... 558

     158 aValTyrThrAspPheTyrGlnAlaPheSerAspIleLeuSerLysMetG 175
     559 ................TACCGTGCA......GGCTTTATCCGTCGTTACG 586

     175 lyGlyTrpLeuLeuProGlyLysAspGlyAsnThrValLysLeuAspVal 191
     587 TCAACTGG.....................CAGCCAAGTCCGTTAGAACAC 615

     192 ThrSerLeuLysAsnAspLeuAsnSerLeuValAsnLysTyrAsnGlnIl 208
     616 ATCGAAATGTTAGAGCAGCTTCGTGTTCTG......TGGTACGGCGAAAA 659

     208 eAsnSerAsnThrValLeuPheProAlaGlnSerGlySerGlyValLysV 225
     660 AATCCATGTTGCTGTT.........GCTCAGGAAGTTCCTGGCACAGGTG 700

     225 alAlaThrGluAlaGluAlaArgGlnTrpLeuSerGluLeuAsnLeuPro 241
     701 TGGATACCCCTGAA.....................GATCTCGACCCGTCG 729

     242 AsnSerCysLeuLysSerTyrGlySerGlyTyrValVal...ThrValAs 257
     730 ACGAATTCCATGGAGATCTACGGTGCTTGCTACTCTATCGAACCGCTGGA 779

     257 pLeuThrProLeuGlnLysMetValGlnAspIleAspGlyLeuGlyAla 273
     780 CCTGCCGCCG.........ATCATTCAGCGTCTGCACGGTCTGTCTGCT 819

seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-609-230A-8

seq_documentation_block:
; Sequence 8, Application US/08609230A
; Patent No. 5866333
;  GENERAL INFORMATION:
;    APPLICANT:  Innerarity, Thomas L.
;    APPLICANT:  Qian, Xiaobing
;    APPLICANT:  Yamanaka, Shinya
;    TITLE OF INVENTION:  Screening Methods to Detect mRNA Targets
;    TITLE OF INVENTION:  of Editing Enzymes
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend and Crew LLP
;      STREET:  Two Embarcadero Center, 8th Floor
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/609,230A
;      FILING DATE:  01-MAR-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fitts, Renee A.
;      REGISTRATION NUMBER:  35,136
;      REFERENCE/DOCKET NUMBER:  02307U-068100US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-326-2400
;      TELEFAX:  650-326-2422
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3751 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  277..2994
US-08-609-230A-8


alignment_scores:
            Quality:   77.50              Length:    314
              Ratio:  0.497                Gaps:     13
 Percent Similarity:  49.682   Percent Identity: 18.471

alignment_block:
 US-09-068-804-7 x US-08-609-230A-8   ..

 Align seg 1/1  to: US-08-609-230A-8  from: 1  to: 3751

       2 LeuAsnIleGlnAsnTyrSerAlaSerProHisProGlyIleValAlaGl 18
    1672 CTTAATGCAGATGAGATTAGTTTGAGGCCTGCTCAGTCGTTTCTAATGAA 1721

      18 uArgProGlnThrProSerAlaSerGluHisValGluThrAlaValValP 35
    1722 TAAAAATCAGGTGCCAAAGCTTCAGCCCCAGATA......ACTATGATTC 1765

      35 roSerThrThrGluHisArgGlyThrAspIleIleSerLeuSerGlnAla 51
    1766 CTCCCAGTGCACAGCCACCACGCACTCAAACACCGCCTCTGGGACAGACA 1815

      52 AlaThrLysIleHisGlnAlaGlnGlnThrLeuGlnSerThrProProIl 68
    1816 .................CCTCAACTTGGTCTCAAAACTAATCCACCACT 1847

      68 eSerGluGluAsnAsnAspGluArgThrLeuAlaArgGlnGlnLeuThrS 85
    1848 TATCCAGGAAAAGCCTGCCAAGACTAGCAAAAAGCAACCACCATCAAAGG 1897

      85 erSerLeuAsnAlaLeuAlaLysSerGlyValSerLeuSerAlaGluGln 101
    1898 AAGAACTACTTAAACTGACCGAAGCCGTTGTGACT.............. 1932

     102 AsnGluAsnLeuArgSerAlaPheSerAlaProThrSerAlaLeuPheSe 118
    1933 ...GACTATCTGAACAGT......GGAAATGCCAACGAGGCTGTCAGTGG 1973

     118 rAlaSerProMetAlaGlnProArgThrThrIleSerAspAlaGluIleT 135
    1974 TGTGAGAGAAATGAGAGCTCCAAAACACTTTCTTCCT.............. 2010

     135 rpAspMetValSerGlnAsnIle........................... 142
    2011 ..GAGATGCTAAGCAAAGTGATCATCCTGTCACTTGATAGAAGCGATGAA 2058

     143 ...SerAlaIleGlyAspSerTyrLeuGlyValTyr...GluAsnValVa 157
```

```
          :::   ::::::|||   :::::::::      ::::::::::::
  2059 GATAAAGAAAAAGCAAGCTCTTTAATCAGTTTACTCAAACAGGAAGCGAT 2108

   157 lAlaValTyrThrAspPheTyrGlnAlaPheSerAspIleLeuSerLysM 174
       :|||:::      :::|||   |||||||||   :::::::|||
  2109 AGCCACAAGTGACAACTTCATGCAGGCTTTCCTGAATGTATTG..... 2151

   174 etGlyGlyTrpLeuLeuProGlyLysAspGlyAsnThrValLysLeuAsp 190
                                    :::      ||||||:::
  2152 ............................GAGCAGTGCCCCAAACTGGAG 2172

   191 ValThrSerLeuLysAsnAspLeuAsnSerLeuValAsnLysTyr...As 206
       |||                    :::|||||||:::   |||   ::
  2173 GTTGAC..................ATCCCCTTGGTGAAATCTTACTTGGC 2204

   206 nGlnIleAsnSerAsnThrValLeuProAlaGlnSerGlySerGlyV 223
       :|||::::::::   :::::::::          |||   |||   :
  2205 ACAGTTTGCAGCTCGTGCTATAATTTCAGAGTTGGTGAGCATTTCCGAAC 2254

   223 alLysValAlaThrGluAla............................ 229
       ::      :::   |||:::
  2255 TAGCTCAACCACTGGAGAGTGGCACCCACTTCCCTCTCTTCTTACTTTGT 2304

   230 .....................GluAlaArgGlnTrpLeuSerGluLeuAs 239
                           :::::::|||:::||||||:::||||||
  2305 CTTCAACAATTAGCTAAATTGCAAGACCGAGAGTGGTTAACCGAACTTTT 2354

   239 nLeuProAsnSerCysLeuLysSerTyrGlySerGlyTyrValValThrV 256
           ::::::
  2355 TCAACAAAGCAAG..................................... 2367

   256 alAspLeuThrProLeuGlnLysMetValGlnAspIleAspGlyLeuGly 272
               :::   :::||||||||||:::::::::||||||
  2368 ........GTCAATATGCAGAAAATGCTGCCAGAAATTGAT......... 2400

   273 AlaProGlyLysAspSerLysLeuGluMetAspAsnAlaLys 286
       ::::::|||||   :::||||||:::   ::::::|||
  2401 ...CAGAATAAGGATCGAATGTTGGAGATTTTGGAAGGAAAG 2439

seq_name: /cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:PCT-US95-14418-3

seq_documentation_block:
; Sequence 3, Application PC/TUS9514418
;  GENERAL INFORMATION:
;    APPLICANT:
;    APPLICANT:
;    TITLE OF INVENTION:  DNA Encoding a Thermostable DNA Polymerase Enzyme
;    NUMBER OF SEQUENCES:  51
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/14418
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gass, David A.
;      REGISTRATION NUMBER:  38,153
;      REFERENCE/DOCKET NUMBER:  28003/32330
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1794 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1794
PCT-US95-14418-3


alignment_scores:
            Quality:   73.50            Length:    198
              Ratio:  0.714              Gaps:     12
 Percent Similarity:  52.020   Percent Identity: 23.232

alignment_block:
 US-09-068-804-7 x PCT-US95-14418-3   ..

 Align seg 1/1  to: PCT-US95-14418-3  from: 1  to: 1794

     40 HisArgGlyThrAspIleIleSerLeuSerGlnAlaAlaThrLysIleHi 56
        |||||||||   |||               ::::::|||:::      ||
    876 CATCGTGGAGAAGAT..............CCTCCAGCACCGGGAGCTCA 910

     56 sGlnAlaGlnGlnThrLeuGlnSerThrProProIleSerGluGluAsnA 73
        ||||||||||:::   |||::::::::::::::::   :::         :
    911 CCAAGCTCAAGAACACCTACGTGGACCCCCTCCCAAGCCTCGTCCACCCG 960

     73 snAsp.GluArgThrLeuAlaArg...GlnGlnLeuThrSerSerLeuAs 88
        ::|||   |||      :::|||   :::|||   :::::::::   ::
    961 AGGACGGGCCGCCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGG 1010

     88 nAlaLeuAlaLysSerGlyValSerLeuSerAlaGluGlnAsnGluAsnL 105
        :   |||:::   |||:::   :::|||         ||||||      :
   1011 GAGGCTTAGTAGCTCCGACCCCAACCTG.........CAGAACATCCCCG 1051

    105 euArgSerAlaPheSerAlaProThrSerAlaLeuPheSerAlaSerPro 121
        ::|||                                      :::|||
   1052 TCCGC.....................................ACCCCC 1062

    122 MetAlaGln......ProArgThrThrIleSerAspAlaGluIleTrpAs 136
        ::::::|||         |||:::   :::::::::|||      |||::
   1063 TTGGGCCAGAGGATCCGCCGGGCCTTCGTGGCCGAGGCGGGA...TGGGC 1109

    136 pMetValSerGlnAsnIleSerAlaIleGly.........AspSerTyrL 150
        :::::|||:::   :::   |||   |||                ::::::|
   1110 GTTGGTGGCCCTGGACTATAGCCAGATAGAGCTCCGCGTCCTCGCCCACC 1159

    150 euGlyValTyrGluAsnValValAlaValTyrThrAspPheTyrGlnAla 166
        ||:::      ||||||::::::   |||            :::|||:::
   1160 TCTCCGGGGACGAGAACCTGATCAGGGTC............TTCCAGGAG 1197

    167 PheSerAspIleLeuSerLysMetGlyGlyTrpLeu.............. 178
              ||||||   ::::::   ::::::|||:::
   1198 GGGAAGGACATCCACACCCAGACCGCAAGCTGGATGTTCGGCGTCCCCCC 1247

    179 ..............LeuProGlyLysAspGlyAsnThrValLysLeuA 190
                        |||      ::::::::::::||||||::::::
   1248 GGAGGCCGTGGACCCCCTGATGCGCCGGGCGGCCAAGACGGTGAACTTCG 1297

    190 spVal.................ThrSerLeuLysAsnAspLeuAsn... 199
        ::|||                        |||   ::::::|||:::
   1298 GCGTCCTCTACGGCATGTCCGCCCATAGGCTCTCCCAGGAGCTTGCCATC 1347

    200 ................SerLeuValAsnLysTyrAsnGln 207
                       ::::::::::::::::|||   |||
   1348 CCCTACGAGGAGGCGGTGGCCTTTATAGAGCGCTACTTCCAA 1389

seq_name: /cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:PCT-US95-15327-3

seq_documentation_block:
; Sequence 3, Application PC/TUS9515327
```

```
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Biologically Active Fragments of
;    TITLE OF INVENTION:  Thermus Flavus DNA Polymerase
;    NUMBER OF SEQUENCES:  51
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/15327
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gass, David A.
;      REGISTRATION NUMBER:  38,153
;      REFERENCE/DOCKET NUMBER:  28003/31716
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1794 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1794
PCT-US95-15327-3


alignment_scores:
            Quality:   73.50              Length:    198
              Ratio:  0.714                Gaps:     12
 Percent Similarity:  52.020    Percent Identity: 23.232

alignment_block:
 US-09-068-804-7 x PCT-US95-15327-3   ..

 Align seq 1/1  to: PCT-US95-15327-3  from: 1  to: 1794

      40 HisArgGlyThrAspIleIleSerLeuSerGlnAlaAlaThrLysIleHi 56
         |||||||||   |||               ::::::|||:::      ||
     876 CATCGTGGAGAAGAT..............CCTCCAGCACCGGGAGCTCA 910

      56 sGlnAlaGlnGlnThrLeuGlnSerThrProProIleSerGluGluAsnA 73
         ||||||||||:::   |||:::::::::::||||||   :::         :
     911 CCAAGCTCAAGAACACCTACGTGGACCCCCTCCCAAGCCTCGTCCACCCG 960

      73 snAsp.GluArgThrLeuAlaArg...GlnGlnLeuThrSerSerLeuAs 88
         ::|||    |||      :::|||   :::|||   :::::::::   ::
     961 AGGACGGGCCGCCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGG 1010

      88 nAlaLeuAlaLysSerGlyValSerLeuSerAlaGluGlnAsnGluAsnL 105
         :   |||:::   |||:::   :::|||         ||||||      :
    1011 GAGGCTTAGTAGCTCCGACCCCAACCTG.........CAGAACATCCCCG 1051

     105 euArgSerAlaPheSerAlaProThrSerAlaLeuPheSerAlaSerPro 121
         ::|||                                    :::|||
    1052 TCCGC.....................................ACCCCC 1062

     122 MetAlaGln......ProArgThrThrIleSerAspAlaGluIleTrpAs 136
         ::::::|||        |||:::   ::::::::::|||      |||::

     1063 TTGGGCCAGAGGATCCGCCGGGCCTTCGTGGCCGAGGCGGGA...TGGGC 1109

      136 pMetValSerGlnAsnIleSerAlaIleGly.........AspSerTyrL 150
          ::::|||:::   :::   |||   |||               ::::::|
     1110 GTTGGTGGCCCTGGACTATAGCCAGATAGAGCTCCGCGTCCTCGCCCACC 1159

      150 euGlyValTyrGluAsnValValAlaValTyrThrAspPheTyrGlnAla 166
          ||:::      ||||||::::::   |||                ::::|||:::
     1160 TCTCCGGGGACGAGAACCTGATCAGGGTC.............TTCCAGGAG 1197

      167 PheSerAspIleLeuSerLysMetGlyGlyTrpLeu.............. 178
              ||||||   ::::::   :::::::|||:::
     1198 GGGAAGGACATCCACACCCAGACCGCAAGCTGGATGTTCGGCGTCCCCCC 1247

      179 ..............LeuProGlyLysAspGlyAsnThrValLysLeuA 190
                        |||      ::::::::::::|||||||:::::::
     1248 GGAGGCCGTGGACCCCCTGATGCGCCGGGCGGCCAAGACGGTGAACTTCG 1297

      190 spVal.................ThrSerLeuLysAsnAspLeuAsn... 199
          ::|||                        |||   ::::::|||:::
     1298 GCGTCCTCTACGGCATGTCCGCCCATAGGCTCTCCCAGGAGCTTGCCATC 1347

      200 .................SerLeuValAsnLysTyrAsnGln 207
                          ::::::::::::::|||   |||
     1348 CCCTACGAGGAGGCGGTGGCCTTTATAGAGCGCTACTTCCAA 1389

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-320-559-1

seq_documentation_block:
; Sequence 1, Application US/08320559
; Patent No. 5633135
;  GENERAL INFORMATION:
;    APPLICANT:  Croce, Carlo
;    APPLICANT:  Canaani, Eli
;    TITLE OF INVENTION:  Diagnostics, Therapeutics and Methods for
;    TITLE OF INVENTION:  Detection and Treatment of Acute Leukemias
;    TITLE OF INVENTION:  Resulting from Chromosome Abnormalities in the
;    TITLE OF INVENTION:  All-1 Region
;    NUMBER OF SEQUENCES:  44
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock Washburn Kurtz Mackiewicz & No. 5633135ris
;      STREET:  One Liberty Place - 46th Floor
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/320,559
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/062,443
;      FILING DATE:  14 MAY 1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/971,094
;      FILING DATE:  30-OCT-92
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/888,830
;      FILING DATE:  27-MAY-92
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/805,093
;      FILING DATE:  11-DEC-91
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DeLuca, Mark
;      REGISTRATION NUMBER:  33,229
;      REFERENCE/DOCKET NUMBER:  TJU-0855
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 568-3100
```

```
;      TELEFAX:  (215) 568-3439
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  14255
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     ANTI-SENSE:  No
US-08-320-559-1


alignment_scores:
            Quality:   82.50              Length:    132
              Ratio:  0.907                Gaps:      4
 Percent Similarity:  68.939   Percent Identity: 21.970

alignment_block:
 US-09-068-804-7 x US-08-320-559-1   ..

 Align seg 1/1  to: US-08-320-559-1  from: 1  to: 14255

      20 ProGlnThrProSerAlaSerGluHisValGluThrAlaValValProSe 36
         :::|||   :::|||:::|||   :::      ::::::   |||   :: :
    1222 TCTCAAATGTCTTCAGACTCCTCTCGATCTAGTAGCCCCAGTGTTGATAC 1271

      36 rThrThrGluHisArgGlyThrAspIleIle...SerLeuSerGlnAlaA 52
         ::::|||:::   ::::::::::::   |||      |||:::::::::
    1272 CTCCACAGACTCTCAGGCTTCTGAGGAGATTCAGGTACTTCCTGAGGAGC 1321

      52 laThrLysIleHisGlnAlaGlnGlnThrLeuGlnSerThrProProIle 68
           :::   :::   ::::::::::::::|||:::   :::::::::
    1322 GGAGCGATACCCCTGAAGTTCATCCTCCACTGCCCATTTCCCAGTCCCCA 1371

      69 SerGluGluAsnAsnAspGluArgThrLeuAlaArgGlnGlnLeuThrSe 85
           :::|||:::|||||   |||      :::|||:::      :::||
    1372 GAAAATGAGAGTAATGATAGGAGA......AGCAGAAGGTATTCAGTGTC 1415

      85 rSerLeuAsnAlaLeuAlaLysSerGlyValSerLeuSerAlaGluGlnA 102
         |      :::      :::::::::         ||||||:::   |||:
    1416 GGAGAGAAGTTTGGATCTAGAACGACGAAAAAATTATCAACTCTACAAA 1465

     102 snGluAsnLeuArgSerAlaPheSerAlaProThrSerAlaLeuPheSer 118
         :::::      :::   :::   |||:::|||:::::::::|||:::::
    1466 GTGCCCCCCAGCAGGAGACCTCCTCGTCTCCACCTCCACCTCTGCTGACT 1515

     119 AlaSerProMetAlaGlnProArgThrThrIleSerAspAlaGluIleTr 135
         ::::::|||      ||||||   ::::::|||||||||          ||
    1516 CCACCGCCACCACTGCAGCCAGCCTCCAGTATCTCTGACCACACACCTTG 1565

     135 pAspMetValSerGlnAsnIleSerAlaIleGlyAspSerTyrLeu 150
         |   :::::::::::::::|||   :::::::::::::::::|||
    1566 G...CTTATGCCTCCAACAATC...CCCTTAGCATCACCATTTTTG 1605

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-327-392-1

seq_documentation_block:
; Sequence 1, Application US/08327392
; Patent No. 5633136
;  GENERAL INFORMATION:
;    APPLICANT:  Croce, Carlo
;    APPLICANT:  Canaani, Eli
;    TITLE OF INVENTION:  ALL-1 Polynucleotides and Monoclonal
;    TITLE OF INVENTION:  Antibodies for Leukemia Detection and
;    TITLE OF INVENTION:  Treatment
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock Washburn Kurtz Mackiewicz & No. 5633136ris
;      STREET:  One Liberty Place - 46th Floor
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
;      COMPUTER:  IBM
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/327,392
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/971,094
;      FILING DATE:  30-OCT-92
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/888,830
;      FILING DATE:  27-MAY-92
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/805,093
;      FILING DATE:  11-DEC-91
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DeLuca, Mark
;      REGISTRATION NUMBER:  33,229
;      REFERENCE/DOCKET NUMBER:  TJU-1331
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 568-3100
;      TELEFAX:  (215) 568-3439
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  14255
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    ANTI-SENSE:  No
US-08-327-392-1


alignment_scores:
            Quality:   82.50              Length:    132
              Ratio:  0.907                Gaps:      4
 Percent Similarity:  68.939   Percent Identity: 21.970

alignment_block:
 US-09-068-804-7 x US-08-327-392-1   ..

 Align seg 1/1  to: US-08-327-392-1  from: 1  to: 14255

      20 ProGlnThrProSerAlaSerGluHisValGluThrAlaValValProSe 36
         :::|||   :::|||:::|||   :::      ::::::   |||   :: :
    1222 TCTCAAATGTCTTCAGACTCCTCTCGATCTAGTAGCCCCAGTGTTGATAC 1271

      36 rThrThrGluHisArgGlyThrAspIleIle...SerLeuSerGlnAlaA 52
         ::::|||:::   ::::::::::::   |||      |||:::::::::
    1272 CTCCACAGACTCTCAGGCTTCTGAGGAGATTCAGGTACTTCCTGAGGAGC 1321

      52 laThrLysIleHisGlnAlaGlnGlnThrLeuGlnSerThrProProIle 68
           :::   :::   ::::::::::::::|||:::   :::::::::
    1322 GGAGCGATACCCCTGAAGTTCATCCTCCACTGCCCATTTCCCAGTCCCCA 1371

      69 SerGluGluAsnAsnAspGluArgThrLeuAlaArgGlnGlnLeuThrSe 85
           :::|||:::|||||   |||      :::|||:::      :::||
    1372 GAAAATGAGAGTAATGATAGGAGA......AGCAGAAGGTATTCAGTGTC 1415

      85 rSerLeuAsnAlaLeuAlaLysSerGlyValSerLeuSerAlaGluGlnA 102
         |      :::      :::::::::         ||||||:::   |||:
    1416 GGAGAGAAGTTTGGATCTAGAACGACGAAAAAATTATCAACTCTACAAA 1465

     102 snGluAsnLeuArgSerAlaPheSerAlaProThrSerAlaLeuPheSer 118
         :::::      :::   :::   |||:::|||:::::::::|||:::::
    1466 GTGCCCCCCAGCAGGAGACCTCCTCGTCTCCACCTCCACCTCTGCTGACT 1515

     119 AlaSerProMetAlaGlnProArgThrThrIleSerAspAlaGluIleTr 135
         ::::::|||      ||||||   ::::::|||||||||          ||
    1516 CCACCGCCACCACTGCAGCCAGCCTCCAGTATCTCTGACCACACACCTTG 1565

     135 pAspMetValSerGlnAsnIleSerAlaIleGlyAspSerTyrLeu 150
```

```
          |   ::::::::::::::::|||   ::::::::::::::::::|||
    1566 G...CTTATGCCTCCAACAATC...CCCTTAGCATCACCATTTTTG 1605

seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-306-691B-55

seq_documentation_block:
; Sequence 55, Application US/08306691B
; Patent No. 5734039
;  GENERAL INFORMATION:
;    APPLICANT:  Calabretta, Bruno
;    APPLICANT:  Skorski, Tomasz
;    TITLE OF INVENTION:  ANTISENSE
;    TITLE OF INVENTION:  OLIGONUCLEOTIDES TARGETING COOPERATING ONCOGENES
;    NUMBER OF SEQUENCES:  55
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Seidel, Gonda, Lavorgna & Monaco, P.C.
;      STREET:  Two Penn Center, Suite 1800
;      CITY:  Philadelphia
;      STATE:  Pennsylvania
;      COUNTRY:  U.S.A.
;      ZIP:  19102
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 720 Kb
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WordPerfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/306,691B
;      FILING DATE:  September 15, 1994
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Monaco, Daniel A.
;      REGISTRATION NUMBER:  30,480
;      REFERENCE/DOCKET NUMBER:  8321-8
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 568-8383
;      TELEFAX:  (215) 568-5549
;      TELEX:  No. 5734039e
;  INFORMATION FOR SEQ ID NO:  55:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  14255 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-306-691B-55


alignment_scores:
            Quality:   82.50            Length:    132
              Ratio:  0.907              Gaps:      4
 Percent Similarity:  68.939   Percent Identity: 21.970

alignment_block:
 US-09-068-804-7 x US-08-306-691B-55   ..

 Align seg 1/1  to: US-08-306-691B-55  from: 1  to: 14255

      20 ProGlnThrProSerAlaSerGluHisValGluThrAlaValValProSe 36
         :::|||   :::||||:::|||   :::      ::::::   |||   ::
    1222 TCTCAAATGTCTTCAGACTCCTCTCGATCTAGTAGCCCCAGTGTTGATAC 1271

      36 rThrThrGluHisArgGlyThrAspIleIle...SerLeuSerGlnAlaA 52
         ::::|||:::   ::::::::::::   |||      |||:::::::::
    1272 CTCCACAGACTCTCAGGCTTCTGAGGAGATTCAGGTACTTCCTGAGGAGC 1321

      52 laThrLysIleHisGlnAlaGlnGlnThrLeuGlnSerThrProProIle 68
         :::   :::   ::::::::::::::::|||:::   :::::::::
    1322 GGAGCGATACCCCTGAAGTTCATCCTCCACTGCCCATTTCCCAGTCCCCA 1371

      69 SerGluGluAsnAsnAspGluArgThrLeuAlaArgGlnGlnLeuThrSe 85

         :::|||:::!!!!!!   |||      :::|||:::      :::||
    1372 GAAAATGAGAGTAATGATAGGAGA......AGCAGAAGGTATTCAGTGTC 1415

      85 rSerLeuAsnAlaLeuAlaLysSerGlyValSerLeuSerAlaGluGlnA 102
         |      :::      :::::::::         ||||||:::   |||:
    1416 GGAGAGAAGTTTTGGATCTAGAACGACGAAAAAATTATCAACTCTACAAA 1465

     102 snGluAsnLeuArgSerAlaPheSerAlaProThrSerAlaLeuPheSer 118
         :::::      :::   :::   |||:::|||:::::::::|||::::::
    1466 GTGCCCCCCAGCAGGAGACCTCCTCGTCTCCACCTCCACCTCTGCTGACT 1515

     119 AlaSerProMetAlaGlnProArgThrThrIleSerAspAlaGluIleTr 135
         ::::::|||      ||||||   ::::::||||||||||         ||
    1516 CCACCGCCACCACTGCAGCCAGCCTCCAGTATCTCTGACCACACACCTTG 1565

     135 pAspMetValSerGlnAsnIleSerAlaIleGlyAspSerTyrLeu 150
         |   ::::::::::::::::|||   ::::::::::::::::::|||
    1566 G...CTTATGCCTCCAACAATC...CCCTTAGCATCACCATTTTTG 1605

seq_name: /cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:PCT-US94-04496-1

seq_documentation_block:
; Sequence 1, Application PC/TUS9404496
;  GENERAL INFORMATION:
;    APPLICANT:  Croce, Carlo
;    APPLICANT:  Canaani, Eli
;    TITLE OF INVENTION:  Diagnostics, Therapeutics and Methods
;    TITLE OF INVENTION:  for Detection and Treatment of Acute Leukemias
;    TITLE OF INVENTION:  Resulting from Chromosome Abnormalities in the All-1
;    NUMBER OF SEQUENCES:  86
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock, Washburn, Kurtz, Mackiewitz &
;      ADDRESSEE:  Norris
;      STREET:  One Liberty Place, 46th floor
;      CITY:  Philadelphia
;      STATE:  Pennsylvania
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/04496
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DeLuca Esq., Mark
;      REGISTRATION NUMBER:  33,229
;      REFERENCE/DOCKET NUMBER:  TJU-1242
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 568-3100
;      TELEFAX:  (215) 568-3439
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  14255
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    ANTI-SENSE:  No
PCT-US94-04496-1


alignment_scores:
            Quality:   82.50            Length:    132
              Ratio:  0.907              Gaps:      4
 Percent Similarity:  68.939   Percent Identity: 21.970

alignment_block:
 US-09-068-804-7 x PCT-US94-04496-1   ..

 Align seg 1/1  to: PCT-US94-04496-1  from: 1  to: 14255
```

```
     20 ProGlnThrProSerAlaSerGluHisValGluThrAlaValValProSe 36
        :::|||   :::|||:::|||   :::      ::::::   |||   ::
   1222 TCTCAAATGTCTTCAGACTCCTCTCGATCTAGTAGCCCCAGTGTTGATAC 1271

     36 rThrThrGluHisArgGlyThrAspIleIle...SerLeuSerGlnAlaA 52
        :::|||:::   ::::::::::::   |||      |||:::::::::::
   1272 CTCCACAGACTCTCAGGCTTCTGAGGAGATTCAGGTACTTCCTGAGGAGC 1321

     52 laThrLysIleHisGlnAlaGlnGlnThrLeuGlnSerThrProProIle 68
          :::   :::   ::::::::::::::|||:::   :::::::::
   1322 GGAGCGATACCCCTGAAGTTCATCCTCCACTGCCCATTTCCCAGTCCCCA 1371

     69 SerGluGluAsnAsnAspGluArgThrLeuAlaArgGlnGlnLeuThrSe 85
        ::|||:::|||||   |||      :::|||:::      :::||
   1372 GAAAATGAGAGTAATGATAGGAGA......AGCAGAAGGTATTCAGTGTC 1415

     85 rSerLeuAsnAlaLeuAlaLysSerGlyValSerLeuSerAlaGluGlnA 102
        |   :::      :::::::::      ||||||:::   |||:
   1416 GGAGAGAAGTTTTGGATCTAGAACGACGAAAAAATTATCAACTCTACAAA 1465

    102 snGluAsnLeuArgSerAlaPheSerAlaProThrSerAlaLeuPheSer 118
        :::::      :::   :::   |||:::|||:::::::::|||:::::
   1466 GTGCCCCCCAGCAGGAGACCTCCTCGTCTCCACCTCCACCTCTGCTGACT 1515

    119 AlaSerProMetAlaGlnProArgThrThrIleSerAspAlaGluIleTr 135
        :::::::|||      ||||||   ::::::|||||||||          ||
   1516 CCACCGCCACCACTGCAGCCAGCCTCCAGTATCTCTGACCACACACCTTG 1565

    135 pAspMetValSerGlnAsnIleSerAlaIleGlyAspSerTyrLeu 150
        |   :::::::::::::::|||   :::::::::::::::::::|||
   1566 G...CTTATGCCTCCAACAATC...CCCTTAGCATCACCATTTTTG 1605

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-977-434-9

seq_documentation_block:
; Sequence 9, Application US/07977434
; Patent No. 5466591
;  GENERAL INFORMATION:
;    APPLICANT:  Gelfand, David H.
;    APPLICANT:  Abramson, Richard D.
;    TITLE OF INVENTION:  5' TO 3' EXONUCLEASE MUTATIONS OF
;    TITLE OF INVENTION:  THERMOSTABLE DNA POLYMERASES
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hoffmann-La Roche Inc.
;      STREET:  340 Kingsland Street
;      CITY:  Nutley
;      STATE:  New Jersey
;      ZIP:  07110-1199
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Macintosh
;      OPERATING SYSTEM:  7
;      SOFTWARE:  WordPerfect 2.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/977,434
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 590,490
;      FILING DATE:  28-SEP-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 590,466
;      FILING DATE:  28-SEP-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 590,213
;      FILING DATE:  28-SEP-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 523,394
;      FILING DATE:  15-MAY-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 143,441
;      FILING DATE:  12-JAN-1988
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 063,509
;      FILING DATE:  17-JUN-1987
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 899,241
;      FILING DATE:  22-AUG-1986
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 746,121
;      FILING DATE:  15-AUG-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/US90/07641
;      FILING DATE:  21-DEC-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 585,471
;      FILING DATE:  20-SEP-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 455,611
;      FILING DATE:  22-DEC-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 609,157
;      FILING DATE:  02-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 557,517
;      FILING DATE:  24-JUL-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Luann Cserr
;      REGISTRATION NUMBER:  31,822
;      REFERENCE/DOCKET NUMBER:  Case No. 5466591 8753
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (510) 814-2972
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2505 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Thermus thermophilus
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..2502
US-07-977-434-9


alignment_scores:
            Quality:   73.50            Length:    198
              Ratio:  0.714              Gaps:     12
 Percent Similarity:  52.020   Percent Identity: 23.232

alignment_block:
 US-09-068-804-7 x US-07-977-434-9   ..

 Align seg 1/1  to: US-07-977-434-9  from: 1  to: 2505

     40 HisArgGlyThrAspIleIleSerLeuSerGlnAlaAlaThrLysIleHi 56
        |||||||||   |||              ::::::|||:::      ||
   1587 CATCGTGGAGAAGAT..............CCTCCAGCACCGGGAGCTCA 1621

     56 sGlnAlaGlnGlnThrLeuGlnSerThrProProIleSerGluGluAsnA 73
        ||||||||||:::   |||::::::::::||||||   :::         :
   1622 CCAAGCTCAAGAACACCTACGTGGACCCCCTCCCAAGCCTCGTCCACCCG 1671

     73 snAsp.GluArgThrLeuAlaArg...GlnGlnLeuThrSerSerLeuAs 88
        ::|||    |||      :::|||   :::|||   :::::::::   ::
   1672 AGGACGGGCCGCCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGG 1721

     88 nAlaLeuAlaLysSerGlyValSerLeuSerAlaGluGlnAsnGluAsnL 105
        :   |||:::   |||:::   :::|||         ||||||      :
   1722 GAGGCTTAGTAGCTCCGACCCCAACCTG.........CAGAACATCCCCG 1762
```

```
     105 euArgSerAlaPheSerAlaProThrSerAlaLeuPheSerAlaSerPro 121
         ::|||                                       :::|||
    1763 TCCGC.....................................ACCCCC 1773

     122 MetAlaGln......ProArgThrThrIleSerAspAlaGluIleTrpAs 136
         :::::::|||         |||:::   ::::::::::|||      |||::
    1774 TTGGGCCAGAGGATCCGCCGGGCCTTCGTGGCCGAGGCGGGT...TGGGC 1820

     136 pMetValSerGlnAsnIleSerAlaIleGly.........AspSerTyrL 150
         ::::|||:::   :::   |||   |||               ::::::|
    1821 GTTGGTGGCCCTGGACTATAGCCAGATAGAGCTCCGCGTCCTCGCCCACC 1870

     150 euGlyValTyrGluAsnValValAlaValTyrThrAspPheTyrGlnAla 166
         ||:::      ||||||::::::   |||             :::|||:::
    1871 TCTCCGGGGACGAAAACCTGATCAGGGTC............TTCCAGGAG 1908

     167 PheSerAspIleLeuSerLysMetGlyGlyTrpLeu.............. 178
             ||||||   ::::::   ::::::|||:::
    1909 GGGAAGGACATCCACACCCAGACCGCAAGCTGGATGTTCGGCGTCCCCCC 1958

     179 ..............LeuProGlyLysAspGlyAsnThrValLysLeuA 190
                        |||      ::::::::::::||||||:::::::
    1959 GGAGGCCGTGGACCCCCTGATGCGCCGGGCGGCCAAGACGGTGAACTTCG 2008

     190 spVal.................ThrSerLeuLysAsnAspLeuAsn... 199
         ::|||                      |||   ::::::|||:::
    2009 GCGTCCTCTACGGCATGTCCGCCCATAGGCTCTCCCAGGAGCTTGCCATC 2058

     200 .................SerLeuValAsnLysTyrAsnGln 207
                         ::::::::::::::|||   |||
    2059 CCCTACGAGGAGGCGGTGGCCTTTATAGAGCGCTACTTCCAA 2100
```

```
seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-458-819-9

seq_documentation_block:
; Sequence 9, Application US/08458819
; Patent No. 5795762
;  GENERAL INFORMATION:
;    APPLICANT:  Gelfand, David H.
;    APPLICANT:  Abramson, Richard D.
;    TITLE OF INVENTION:  5' TO 3' EXONUCLEASE MUTATIONS OF
;    TITLE OF INVENTION:  THERMOSTABLE DNA POLYMERASES
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hoffmann-La Roche Inc.
;      STREET:  340 Kingsland Street
;      CITY:  Nutley
;      STATE:  New Jersey
;      ZIP:  07110-1199
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Macintosh
;      OPERATING SYSTEM:  7
;      SOFTWARE:  WordPerfect 2.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/458,819
;      FILING DATE:  02-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/977,434
;      FILING DATE:  23-FEB-1993
;      APPLICATION NUMBER:  US 590,490
;      FILING DATE:  28-SEP-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 590,466
;      FILING DATE:  28-SEP-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 590,213
;      FILING DATE:  28-SEP-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 523,394
;      FILING DATE:  15-MAY-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 143,441
;      FILING DATE:  12-JAN-1988
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 063,509
;      FILING DATE:  17-JUN-1987
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 899,241
;      FILING DATE:  22-AUG-1986
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 746,121
;      FILING DATE:  15-AUG-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/US90/07641
;      FILING DATE:  21-DEC-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 585,471
;      FILING DATE:  20-SEP-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 455,611
;      FILING DATE:  22-DEC-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 609,157
;      FILING DATE:  02-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 557,517
;      FILING DATE:  24-JUL-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Luann Cserr
;      REGISTRATION NUMBER:  31,822
;      REFERENCE/DOCKET NUMBER:  Case No. 5795762 8753
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (510) 814-2972
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2505 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Thermus thermophilus
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..2502
US-08-458-819-9


alignment_scores:
           Quality:   73.50              Length:    198
             Ratio:  0.714                Gaps:     12
 Percent Similarity:  52.020   Percent Identity: 23.232

alignment_block:
 US-09-068-804-7 x US-08-458-819-9    ..

 Align seg 1/1  to: US-08-458-819-9  from: 1  to: 2505

      40 HisArgGlyThrAspIleIleSerLeuSerGlnAlaAlaThrLysIleHi 56
         |||||||||   |||               ::::::|||:::      ||
    1587 CATCGTGGAGAAGAT.............CCTCCAGCACCGGGAGCTCA 1621

      56 sGlnAlaGlnGlnThrLeuGlnSerThrProProIleSerGluGluAsnA 73
         ||||||||||:::   |||:::::::::|||||   :::      :
    1622 CCAAGCTCAAGAACACCTACGTGGACCCCCTCCCAAGCCTCGTCCACCCG 1671

      73 snAsp.GluArgThrLeuAlaArg...GlnGlnLeuThrSerSerLeuAs 88
         ::|||    |||      :::|||   :::|||   :::::::::   ::
    1672 AGGACGGGCCGCCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGG 1721

      88 nAlaLeuAlaLysSerGlyValSerLeuSerAlaGluGlnAsnGluAsnL 105
```

```
      :   |||:::   |||:::   :::|||         ||||||      :
   1722 GAGGCTTAGTAGCTCCGACCCCAACCTG.........CAGAACATCCCCG 1762

    105 euArgSerAlaPheSerAlaProThrSerAlaLeuPheSerAlaSerPro 121
        ::|||                                         :::|||
   1763 TCCGC.....................................ACCCCC 1773

    122 MetAlaGln......ProArgThrThrIleSerAspAlaGluIleTrpAs 136
        ::::::|||        |||:::   :::::::::|||      |||::
   1774 TTGGGCCAGAGGATCCGCCGGGCCTTCGTGGCCGAGGCGGGT...TGGGC 1820

    136 pMetValSerGlnAsnIleSerAlaIleGly.........AspSerTyrL 150
        :::::|||:::   :::   |||   |||               :::::::|
   1821 GTTGGTGGCCCTGGACTATAGCCAGATAGAGCTCCGCGTCCTCGCCCACC 1870

    150 euGlyValTyrGluAsnValValAlaValTyrThrAspPheTyrGlnAla 166
        ||:::      ||||||::::::   |||             :::|||:::
   1871 TCTCCGGGGACGAAAACCTGATCAGGGTC............TTCCAGGAG 1908

    167 PheSerAspIleLeuSerLysMetGlyGlyTrpLeu.............. 178
              ||||||   ::::::   :::::::|||:::
   1909 GGGAAGGACATCCACACCCAGACCGCAAGCTGGATGTTCGGCGTCCCCCC 1958

    179 ................LeuProGlyLysAspGlyAsnThrValLysLeuA 190
                        |||      ::::::::::::|||||||:::::::
   1959 GGAGGCCGTGGACCCCCTGATGCGCCGGGCGGCCAAGACGGTGAACTTCG 2008

    190 spVal..................ThrSerLeuLysAsnAspLeuAsn... 199
        ::|||                          |||   :::::::|||:::
   2009 GCGTCCTCTACGGCATGTCCGCCCATAGGCTCTCCCAGGAGCTTGCCATC 2058

    200 ..................SerLeuValAsnLysTyrAsnGln 207
                           ::::::::::::::|||   |||
   2059 CCCTACGAGGAGGCGGTGGCCTTTATAGAGCGCTACTTCCAA 2100
```

seq_name: /cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:PCT-US91-07035-9

```
seq_documentation_block:
; Sequence 9, Application PC/TUS9107035
;  GENERAL INFORMATION:
;    APPLICANT:  Gelfand, David H.
;    APPLICANT:  Abramson, Richard D.
;    TITLE OF INVENTION:  5' TO 3' EXONUCLEASE MUTATIONS OF
;    TITLE OF INVENTION:  THERMOSTABLE DNA POLYMERASES
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cetus Corporation
;      STREET:  1400 Fifty-third Street
;      CITY:  Emeryville
;      STATE:  California
;      ZIP:  94608
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WordPerfect 5.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US91/07035
;      FILING DATE:  19910930
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 590,490
;      FILING DATE:  28-SEP-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 590,466
;      FILING DATE:  28-SEP-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 590,213
;      FILING DATE:  28-SEP-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 523,394
;      FILING DATE:  15-MAY-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 143,441
;      FILING DATE:  12-JAN-1988
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 063,509
;      FILING DATE:  17-JUN-1987
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 899,241
;      FILING DATE:  22-AUG-1986
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 746,121
;      FILING DATE:  15-AUG-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/US90/07641
;      FILING DATE:  21-DEC-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 585,471
;      FILING DATE:  20-SEP-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 455,611
;      FILING DATE:  22-DEC-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 609,157
;      FILING DATE:  02-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 557,517
;      FILING DATE:  24-JUL-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sias Ph.D, Stacey R.
;      REGISTRATION NUMBER:  32,630
;      REFERENCE/DOCKET NUMBER:  Case No. 2580
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-420-3300
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2505 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Thermus thermophilus
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..2502
PCT-US91-07035-9


alignment_scores:
            Quality:   73.50              Length:    198
              Ratio:  0.714                Gaps:     12
 Percent Similarity:  52.020   Percent Identity: 23.232

alignment_block:
 US-09-068-804-7 x PCT-US91-07035-9    ..

 Align seg 1/1  to: PCT-US91-07035-9  from: 1  to: 2505

      40 HisArgGlyThrAspIleIleSerLeuSerGlnAlaAlaThrLysIleHi 56
         |||||||||   |||             :::::::|||:::       ||
    1587 CATCGTGGAGAAGAT..............CCTCCAGCACCGGGAGCTCA 1621

      56 sGlnAlaGlnGlnThrLeuGlnSerThrProProIleSerGluGluAsnA 73
         ||||||||||:::   |||::::::::::||||||   :::         :
    1622 CCAAGCTCAAGAACACCTACGTGGACCCCCTCCCAAGCCTCGTCCACCCG 1671

      73 snAsp.GluArgThrLeuAlaArg...GlnGlnLeuThrSerSerLeuAs 88
         ::|||    |||      :::|||   :::|||   :::::::::   ::
    1672 AGGACGGGCCGCCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGG 1721

      88 nAlaLeuAlaLysSerGlyValSerLeuSerAlaGluGlnAsnGluAsnL 105
         :   |||:::   |||:::   :::|||         ||||||      :
```

```
    1722 GAGGCTTAGTAGCTCCGACCCCAACCTG.........CAGAACATCCCCG 1762

     105 euArgSerAlaPheSerAlaProThrSerAlaLeuPheSerAlaSerPro 121
         ::|||                                       :::|||
    1763 TCCGC.....................................ACCCCC 1773

     122 MetAlaGln......ProArgThrThrIleSerAspAlaGluIleTrpAs 136
         :::::::|||         |||:::   :::::::::|||      |||::
    1774 TTGGGCCAGAGGATCCGCCGGGCCTTCGTGGCCGAGGCGGGT...TGGGC 1820

     136 pMetValSerGlnAsnIleSerAlaIleGly.........AspSerTyrL 150
         ::::|||:::   :::   |||   |||                 ::::::|
    1821 GTTGGTGGCCCTGGACTATAGCCAGATAGAGCTCCGCGTCCTCGCCCACC 1870

     150 euGlyValTyrGluAsnValValAlaValTyrThrAspPheTyrGlnAla 166
         ||:::      ||||||:::::   |||            :::|||:::
    1871 TCTCCGGGGACGAAAACCTGATCAGGGTC............TTCCAGGAG 1908

     167 PheSerAspIleLeuSerLysMetGlyGlyTrpLeu.............. 178
             ||||||   ::::::   :::::::|||:::
    1909 GGGAAGGACATCCACACCCAGACCGCAAGCTGGATGTTCGGCGTCCCCCC 1958

     179 ...............LeuProGlyLysAspGlyAsnThrValLysLeuA 190
                       |||      ::::::::::::|||||:::::::
    1959 GGAGGCCGTGGACCCCCTGATGCGCCGGGCGGCCAAGACGGTGAACTTCG 2008

     190 spVal.................ThrSerLeuLysAsnAspLeuAsn... 199
         ::|||                     |||   :::::::|||:::
    2009 GCGTCCTCTACGGCATGTCCGCCCATAGGCTCTCCCAGGAGCTTGCCATC 2058

     200 .................SerLeuValAsnLysTyrAsnGln 207
                         ::::::::::::::|||   |||
    2059 CCCTACGAGGAGGCGGTGGCCTTTATAGAGCGCTACTTCCAA 2100
```

seq_name: /cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-648-657-14

```
seq_documentation_block:
; Sequence 14, Application US/08648657
; Patent No. 5885813
;  GENERAL INFORMATION:
;    APPLICANT:  Davis, Maria
;    APPLICANT:  Moffett, R. Bruce
;    APPLICANT:  Fuller, Carl W.
;    TITLE OF INVENTION:  THERMOSTABLE DNA
;    TITLE OF INVENTION:  POLYMERASES
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071-2066
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  IBM P.C. DOS 5.0
;      SOFTWARE:  Word Perfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/648,657
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;    PRIOR APPLICATION DATA:  including application
;    PRIOR APPLICATION DATA:  described below:          one
;      APPLICATION NUMBER:  US 08/455,686
;      FILING DATE:  May 31, 1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warburg, Richard J.
;      REGISTRATION NUMBER:  32,327
;      REFERENCE/DOCKET NUMBER:  219/304
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1686 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  FY4
;      LOCATION:  1...1683
US-08-648-657-14


alignment_scores:
            Quality:   71.50              Length:    198
              Ratio:  0.694                Gaps:     12
 Percent Similarity:  52.020   Percent Identity: 22.727

alignment_block:
 US-09-068-804-7 x US-08-648-657-14   ..

 Align seg 1/1  to: US-08-648-657-14  from: 1  to: 1686

      40 HisArgGlyThrAspIleIleSerLeuSerGlnAlaAlaThrLysIleHi 56
         ||||||||   |||                 ::::::|||:::     ||
     768 CATCGTGGAGAAGAT..............CCTCCAGCACCGGGAGCTCA 802

      56 sGlnAlaGlnGlnThrLeuGlnSerThrProProIleSerGluGluAsnA 73
         ||||||||||:::   |||:::::::::|||||   :::        :
     803 CCAAGCTCAAGAACACCTACGTGGACCCCCTCCCAAGCCTCGTCCACCCG 852

      73 snAsp.GluArgThrLeuAlaArg...GlnGlnLeuThrSerSerLeuAs 88
         ::|||   |||      :::|||   :::|||   :::::::::   ::
     853 AGGACGGGCCGCCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGG 902

      88 nAlaLeuAlaLysSerGlyValSerLeuSerAlaGluGlnAsnGluAsnL 105
         :   |||:::   |||:::   :::|||         ||||||      :
     903 GAGGCTTAGTAGCTCCGACCCCAACCTG.........CAGAACATCCCCG 943

     105 euArgSerAlaPheSerAlaProThrSerAlaLeuPheSerAlaSerPro 121
         ::|||                                       :::|||
     944 TCCGC.....................................ACCCCC 954

     122 MetAlaGln......ProArgThrThrIleSerAspAlaGluIleTrpAs 136
         :::::::|||         |||:::   :::::::::|||      |||::
     955 TTGGGCCAGAGGATCCGCCGGGCCTTCGTGGCCGAGGCGGGT...TGGGC 1001

     136 pMetValSerGlnAsnIleSerAlaIleGly.........AspSerTyrL 150
         ::::|||:::   :::   |||   |||                 ::::::|
    1002 GTTGGTGGCCCTGGACTATAGCCAGATAGAGCTCCGCGTCCTCGCCCACC 1051

     150 euGlyValTyrGluAsnValValAlaValTyrThrAspPheTyrGlnAla 166
         ||:::      ||||||:::::   |||            :::|||:::
    1052 TCTCCGGGGACGAAAACCTGATCAGGGTC............TTCCAGGAG 1089

     167 PheSerAspIleLeuSerLysMetGlyGlyTrpLeu.............. 178
             ||||||   ::::::   :::::::|||:::
    1090 GGGAAGGACATCCACACCCAGACCGCAAGCTGGATGTTCGGCGTCCCCCC 1139

     179 ...............LeuProGlyLysAspGlyAsnThrValLys.... 188
                       |||      ::::::::::::|||||||:::
    1140 GGAGGCCGTGGACCCCCTGATGCGCCGGGCGGCCAAGACGGTGAACTACG 1189

     189 .............LeuAspValThrSerLeuLysAsnAspLeuAsn... 199
                     :::::::::      |||   :::::|||:::
    1190 GCGTCCTCTACGGCATGTCCGCCCATAGGCTCTCCCAGGAGCTAGCCATC 1239

     200 .................SerLeuValAsnLysTyrAsnGln 207
                         ::::::::::::::|||   |||
    1240 CCCTACGAAGAAGCGGTGGCCTTTATAGAGCGCTACTTCCAA 1281
```

```
seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-384-490-30

seq_documentation_block:
; Sequence 30, Application US/08384490
; Patent No. 5618711
;  GENERAL INFORMATION:
;    APPLICANT:  Gelfand, David H.
;    APPLICANT:  Lawyer, Frances C.
;    APPLICANT:  Stoffel, Susanne
;    TITLE OF INVENTION:  Recombinant Expression Vectors and
;    TITLE OF INVENTION:  Purification Methods for Thermus Thermophilus DNA
;    TITLE OF INVENTION:  Polymerase
;    NUMBER OF SEQUENCES:  31
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hoffmann-La Roche Inc.
;      STREET:  340 Kingsland Street
;      CITY:  Nutley
;      STATE:  New Jersey
;      COUNTRY:  U.S.A.
;      ZIP:  07110
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/384,490
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/148,133
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sias, Stacey R.
;      REGISTRATION NUMBER:  32,630
;      REFERENCE/DOCKET NUMBER:  8887
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (510) 814-2863
;      TELEFAX:  (510) 814-2977
;  INFORMATION FOR SEQ ID NO:  30:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2640 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-384-490-30


alignment_scores:
            Quality:   73.50            Length:    198
              Ratio:  0.714              Gaps:     12
 Percent Similarity:  52.020   Percent Identity: 23.232

alignment_block:
 US-09-068-804-7 x US-08-384-490-30   ..

 Align seg 1/1  to: US-08-384-490-30  from: 1  to: 2640

      40 HisArgGlyThrAspIleIleSerLeuSerGlnAlaAlaThrLysIleHi 56
         |||||||||   |||              :::::|||:::      ||
    1707 CATCGTGGAGAAGAT.............CCTCCAGCACCGGGAGCTCA 1741

      56 sGlnAlaGlnGlnThrLeuGlnSerThrProProIleSerGluGluAsnA 73
         ||||||||||:::   |||::::::::::||||||   :::         :
    1742 CCAAGCTCAAGAACACCTACGTGGACCCCCTCCCAAGCCTCGTCCACCCG 1791

      73 snAsp.GluArgThrLeuAlaArg...GlnGlnLeuThrSerSerLeuAs 88
         ::|||    |||      :::|||   :::|||   :::::::::   ::
    1792 AGGACGGGCCGCCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGG 1841

      88 nAlaLeuAlaLysSerGlyValSerLeuSerAlaGluGlnAsnGluAsnL 105
         :   |||:::   |||:::   :::|||         ||||||      :
    1842 GAGGCTTAGTAGCTCCGACCCCAACCTG.........CAGAACATCCCCG 1882

     105 euArgSerAlaPheSerAlaProThrSerAlaLeuPheSerAlaSerPro 121
         ::|||                                        :::|||
    1883 TCCGC.....................................ACCCCC 1893

     122 MetAlaGln......ProArgThrThrIleSerAspAlaGluIleTrpAs 136
         :::::|||         |||:::   :::::::::|||      |||::
    1894 TTGGGCCAGAGGATCCGCCGGGCCTTCGTGGCCGAGGCGGGT...TGGGC 1940

     136 pMetValSerGlnAsnIleSerAlaIleGly.........AspSerTyrL 150
         ::::|||:::   :::   |||   |||                :::::|
    1941 GTTGGTGGCCCTGGACTATAGCCAGATAGAGCTCCGCGTCCTCGCCCACC 1990

     150 euGlyValTyrGluAsnValValAlaValTyrThrAspPheTyrGlnAla 166
         ||:::      ||||||::::::   |||            :::|||:::
    1991 TCTCCGGGGACGAAAACCTGATCAGGGTC...........TTCCAGGAG 2028

     167 PheSerAspIleLeuSerLysMetGlyGlyTrpLeu.............. 178
               ||||||   ::::::   ::::::|||:::
    2029 GGGAAGGACATCCACACCCAGACCGCAAGCTGGATGTTCGGCGTCCCCCC 2078

     179 ..............LeuProGlyLysAspGlyAsnThrValLysLeuA 190
                         |||      :::::::::::::||||||:::::::
    2079 GGAGGCCGTGGACCCCCTGATGCGCCGGGCGGCCAAGACGGTGAACTTCG 2128

     190 spVal.................ThrSerLeuLysAsnAspLeuAsn... 199
         ::|||                          |||   ::::::|||:::
    2129 GCGTCCTCTACGGCATGTCCGCCCATAGGCTCTCCCAGGAGCTTGCCATC 2178

     200 .................SerLeuValAsnLysTyrAsnGln 207
                          ::::::::::::::::|||   |||
    2179 CCCTACGAGGAGGCGGTGGCCTTTATAGAGCGCTACTTCCAA 2220

seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-459-383-30

seq_documentation_block:
; Sequence 30, Application US/08459383
; Patent No. 5741690
;  GENERAL INFORMATION:
;    APPLICANT:  Gelfand, David H.
;    APPLICANT:  Lawyer, Frances C.
;    APPLICANT:  Stoffel, Susanne
;    TITLE OF INVENTION:  Recombinant Expression Vectors and
;    TITLE OF INVENTION:  Purification Methods for Thermus Thermophilus DNA
;    TITLE OF INVENTION:  Polymerase
;    NUMBER OF SEQUENCES:  31
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hoffmann-La Roche Inc.
;      STREET:  340 Kingsland Street
;      CITY:  Nutley
;      STATE:  New Jersey
;      COUNTRY:  U.S.A.
;      ZIP:  07110
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/459,383
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/384,490
;      FILING DATE:
;      APPLICATION NUMBER:  US/08/148,133
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sias, Stacey R.
;      REGISTRATION NUMBER:  32,630
;      REFERENCE/DOCKET NUMBER:  8887
```

```
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (510) 814-2863
;      TELEFAX:  (510) 814-2977
;  INFORMATION FOR SEQ ID NO:  30:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2640 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-459-383-30


alignment_scores:
            Quality:   73.50              Length:    198
              Ratio:  0.714                Gaps:     12
 Percent Similarity:  52.020   Percent Identity: 23.232

alignment_block:
 US-09-068-804-7 x US-08-459-383-30   ..

 Align seg 1/1  to: US-08-459-383-30  from: 1  to: 2640

      40 HisArgGlyThrAspIleIleSerLeuSerGlnAlaAlaThrLysIleHi 56
         |||||||||   |||                :::::::|||:::      ||
    1707 CATCGTGGAGAAGAT..............CCTCCAGCACCGGGAGCTCA 1741

      56 sGlnAlaGlnGlnThrLeuGlnSerThrProProIleSerGluGluAsnA 73
         ||||||||||:::   |||::::::::::||||||   :::         :
    1742 CCAAGCTCAAGAACACCTACGTGGACCCCCTCCCAAGCCTCGTCCACCCG 1791

      73 snAsp.GluArgThrLeuAlaArg...GlnGlnLeuThrSerSerLeuAs 88
         ::|||    |||       :::|||   :::|||   :::::::::   ::
    1792 AGGACGGGCCGCCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGG 1841

      88 nAlaLeuAlaLysSerGlyValSerLeuSerAlaGluGlnAsnGluAsnL 105
         :   |||:::   |||:::   :::|||         ||||||      :
    1842 GAGGCTTAGTAGCTCCGACCCCAACCTG.........CAGAACATCCCCG 1882

     105 euArgSerAlaPheSerAlaProThrSerAlaLeuPheSerAlaSerPro 121
         ::|||                                       :::|||
    1883 TCCGC.....................................ACCCCC 1893

     122 MetAlaGln......ProArgThrThrIleSerAspAlaGluIleTrpAs 136
         :::::::|||         |||:::   :::::::::|||      |||::
    1894 TTGGGCCAGAGGATCCGCCGGGCCTTCGTGGCCGAGGCGGGT...TGGGC 1940

     136 pMetValSerGlnAsnIleSerAlaIleGly.........AspSerTyrL 150
         ::::|||:::   :::   |||   |||               ::::::|
    1941 GTTGGTGGCCCTGGACTATAGCCAGATAGAGCTCCGCGTCCTCGCCCACC 1990

     150 euGlyValTyrGluAsnValValAlaValTyrThrAspPheTyrGlnAla 166
         ||:::      ||||||::::::   |||            :::|||:::
    1991 TCTCCGGGGACGAAAACCTGATCAGGGTC............TTCCAGGAG 2028

     167 PheSerAspIleLeuSerLysMetGlyGlyTrpLeu.............. 178
               ||||||   ::::::   ::::::|||:::
    2029 GGGAAGGACATCCACACCCAGACCGCAAGCTGGATGTTCGGCGTCCCCCC 2078

     179 ...............LeuProGlyLysAspGlyAsnThrValLysLeuA 190
                         |||      ::::::::::::||||||:::::::
    2079 GGAGGCCGTGGACCCCCTGATGCGCCGGGCGGCCAAGACGGTGAACTTCG 2128

     190 spVal.................ThrSerLeuLysAsnAspLeuAsn... 199
         ::|||                        |||   ::::::|||:::
    2129 GCGTCCTCTACGGCATGTCCGCCCATAGGCTCTCCCAGGAGCTTGCCATC 2178

     200 .................SerLeuValAsnLysTyrAsnGln 207
                          ::::::::::::::|||   |||
    2179 CCCTACGAGGAGGCGGTGGCCTTTATAGAGCGCTACTTCCAA 2220
```

This Page Blank (uspto)

OM of: US-09-068-804-7 to: N_Geneseq_36:* out_format : pfs

Date: Oct 5, 1999 8:04 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=x1p
-Q=/cgn2_1/USPTO_spool/US09068804/runat_05101999_145135_10341/app_query.fasta.1
-DB=N_Geneseq_36 -QFMT=fastap -SUFFIX=rng -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=pam150
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=escore
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09068804
-NCPU=6 -ICPU=3 -WAIT -THREADS=1

Search information block:
Query: US-09-068-804-7
Query length: 344
Database: N_Geneseq_36:*
Database sequences: 311585
Database length: 125096042
Search time (sec): 156.120000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ..
N_Geneseq_36:T67030	+	1477.00	3108.17	1.6e-165	1032	! Salmonella secreted protein Ss
N_Geneseq_36:T67035	+	1445.00	3027.69	4.8e-161	3441	! Salmonella virulence factor Ss
N_Geneseq_36:X39703	+	77.50	144.29	1.95	744	! Gastric cancer associated gene.
N_Geneseq_36:X20263	-	85.50	135.21	6.25	8905	! Borrelia burgdorferi polynucle
N_Geneseq_36:T14053	+	79.50	134.26	7.06	2899	! Fungal signal recognition part
N_Geneseq_36:V60373	+	75.00	132.55	8.79	1374	! Human Di12 protein coding sequ
N_Geneseq_36:X09011	+	74.50	132.35	9.02	1266	! Brn-3a polynucleotide. New pol
N_Geneseq_36:T68058	+	74.00	131.72	9.78	1215	! H. pylori cytoplasmic protein
N_Geneseq_36:T68849	+	80.50	131.31	10.30	4698	! Photorhabdus luminescens insec
N_Geneseq_36:X12983	+	78.00	130.14	11.97	3169	! Enterococcus faecalis genome c
N_Geneseq_36:V71025	+	75.00	129.58	12.87	1824	! p38-green flourescent protein
N_Geneseq_36:X09010	+	73.00	129.12	13.65	1272	! Brn-3a polynucleotide. New pol
N_Geneseq_36:V71180	+	75.50	128.95	13.95	2143	! Clone 1711346IH, the PS108 gen
N_Geneseq_36:V71181	+	75.50	128.90	14.03	2152	! Consensus sequence of the PS10
N_Geneseq_36:V29926	+	79.50	128.90	14.04	4832	! tccB gene from the tcc genomic
N_Geneseq_36:Q38241	+	72.50	128.86	14.12	1179	! HCV CKS-NS5I recombinant antig
N_Geneseq_36:Q38256	+	72.50	128.86	14.12	1179	! HCV CKS-NS5I recombinant antig
N_Geneseq_36:Q38271	+	72.50	128.86	14.12	1179	! HCV CKS-NS5I fusion gene in pH
N_Geneseq_36:Q21683	+	72.50	128.83	14.17	1182	! pHCV-49 - recombinant fusion p
N_Geneseq_36:N92334	+	70.50	128.74	14.33	796	! Sequence of a ribosome binding
N_Geneseq_36:T13668	+	75.50	127.64	16.50	2427	! AcNPV ORF 66, residues 55292-5
N_Geneseq_36:Q24328	+	73.50	127.55	16.70	1635	! Mutant thermostable DNA polyme
N_Geneseq_36:N82410	+	72.50	127.28	17.28	1370	! Plasmid pRMIT encoding 70 kD m
N_Geneseq_36:T97955	+	78.50	127.15	17.58	4665	! Human epididymis-specific rece
N_Geneseq_36:Q54431	+	70.50	126.64	18.76	972	! Barley lectin cDNA clone BLc3.
N_Geneseq_36:T32327	+	73.50	126.57	18.92	1794	! Thermus flavus DNA polymerase
N_Geneseq_36:T27255	+	73.50	126.57	18.92	1794	! DNA polymerase I exonuclease-f
N_Geneseq_36:Q24327	+	73.50	125.98	20.43	1899	! Mutant thermostable DNA polyme
N_Geneseq_36:Q45339	+	82.50	125.80	20.90	11907	! cDNA encoding htrx. Nucleic a
N_Geneseq_36:Q24322	+	72.50	125.43	21.92	1635	! Mutant thermostable DNA polyme
N_Geneseq_36:Q24326	+	73.50	125.21	22.54	2043	! Mutant thermostable DNA polyme
N_Geneseq_36:V21388	+	69.50	124.53	24.58	971	! Fibroblast growth factor homol
N_Geneseq_36:V58586	+	75.50	124.08	26.07	3410	! Prostate tumour specific gene
N_Geneseq_36:V61201	+	75.50	124.08	26.07	3410	! Full length cDNA sequence of p
N_Geneseq_36:Q24325	+	73.50	124.07	26.08	2277	! Mutant thermostable DNA polyme
N_Geneseq_36:Q43526	+	82.50	123.91	26.63	14255	! ALL-1 gene cDNA. Detection an
N_Geneseq_36:Q75181	+	82.50	123.91	26.63	14255	! ALL-1 (acute lymphocytic leuk
N_Geneseq_36:V20477	+	82.50	123.91	26.63	14255	! Human ALL-1/AF-4 oncogene. An
N_Geneseq_36:Q24321	+	72.50	123.86	26.81	1899	! Mutant thermostable DNA polyme
N_Geneseq_36:N81539	+	67.50	123.83	26.90	693	! B2-R1 fragment contg. region en
N_Geneseq_36:Q24324	+	73.50	123.65	27.52	2370	! Mutant thermostable DNA polyme
N_Geneseq_36:V71177	+	63.00	123.29	28.84	294	! PS108 gene-specific clone 37053
N_Geneseq_36:X23749	-	71.00	123.16	29.30	1498	! Rice anthranilate synthase
N_Geneseq_36:Q24320	+	72.50	123.09	29.58	2043	! Mutant thermostable DNA pol
N_Geneseq_36:Q24323	+	73.50	123.07	29.65	2505	! Mutant thermostable DNA pol

seq_name: N_Geneseq_36:T67030

seq_documentation_block:
ID T67030 standard; DNA; 1032 BP.
AC T67030;
DT 04-AUG-1997 (first entry)
DE Salmonella secreted protein SspD DNA.
KW Salmonella secreted protein; SspD; bacterial-mediated endocytosis;
KW diagnosis; therapy; vaccine; attenuation; virulence; ss.
OS Salmonella typhimurium.
FH Key Location/Qualifiers
FT cds 1. .1032
FT /*tag= a
FT /product= SspD
FT /note= "the SspD amino acid sequence deduced
FT from the coding sequence does not fully
FT correspond to the SspD sequence given in
FT Fig 14"
PN WO9718225-A1.
PD 22-MAY-1997.
PF 14-NOV-1996; U18504.
PR 14-NOV-1995; US-006733.
PA (GEHO) GEN HOSPITAL CORP.
PI Miller SI;
DR WPI; 97-289217/26.
DR P-PSDB; W15270.
PT New isolated Salmonella secreted proteins and related genes - used
PT to develop products for the detection, treatment or prevention of
PT Salmonella infections
PS Claim 7; Fig 21; 95pp; English.
CC A DNA molecule (T67030) codes for Salmonella secreted protein SspD
CC (W15270), a virulence factor involved in bacterial mediated
CC endocytosis, which is able to translocate into human epithelial
CC cells. To identify Ssp genes, S. typhimurium mutants were
CC investigated in invasion assays on HEp-2 epithelial cells. 2
CC Ssp mutants with transposon insertions located between spaT and
CC prgH were identified. Of these, one mutant lacked Ssp of 87,
CC 42 and 36 kDa. The invasiveness of this mutant implicated the
CC 42 and 36 kDa Ssp in invasion. DNA sequencing of this region (see
CC also T67035) identified 2 complete and 2 partial open reading
CC frames, designated sspB, sspC (T67029), sspD (T67030) and sspA
CC (T67031). These nucleic acids can be used in methods for diagnosing
CC Salmonella infection and in methods of targetting an antigen to
CC epithelial cells and of inducing a cytotoxic T cell immune response
CC in a mammal. Ssp gene mutation can be used to create attenuated
CC Salmonella strains useful as live vaccines.
SQ Sequence 1032 BP; 304 A; 242 C; 246 G; 240 T;

alignment_scores:
Quality: 1477.00 Length: 343
Ratio: 4.306 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.708

alignment_block:
US-09-068-804-7 x T67030 ..

Align seg 1/1 to: T67030 from: 1 to: 1032

```
      1 MetLeuAsnIleGlnAsnTyrSerAlaSerProHisProGlyIleValAl 17
        |||||||||||||||||||||||||||||||||||||||||||||||||||
      1 ATGCTTAATATTCAAAATTATTCCGCTTCTCCTCATCCGGGGATCGTTGC 50

     17 aGluArgProGlnThrProSerAlaSerGluHisValGluThrAlaValV 34
        |||||||||||||||||||||||||||||||||||||||||||||||||||
     51 CGAACGGCCGCAGACTCCCTCGGCGAGCGAGCACGTCGAGACTGCCGTGG 100

     34 alProSerThrThrGluHisArgGlyThrAspIleIleSerLeuSerGln 50
        |||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
  101 TACCGTCTACCACAGAACATCGCGGTACAGATATCATTTCATTATCGCAG 150

   51 AlaAlaThrLysIleHisGlnAlaGlnGlnThrLeuGlnSerThrProPr 67
      |||||||||||||||||||||||||||||||||||||||||||||||||||
  151 GCGGCTACTAAAATCCACCAGGCACAGCAGACGCTGCAGTCAACGCCACC 200

   67 oIleSerGluGluAsnAsnAspGluArgThrLeuAlaArgGlnGlnLeuT 84
      |||||||||||||||||||||||||||||||||||||||||||||||||||
  201 GATCTCTGAAGAGAATAATGACGAGCGCACGCTGGCGCGCCAGCAGTTGA 250

   84 hrSerSerLeuAsnAlaLeuAlaLysSerGlyValSerLeuSerAlaGlu 100
      |||||||||||||||||||||||||||||||||||||||||||||||||||
  251 CCAGCAGCCTGAATGCGCTGGCGAAGTCCGGCGTGTCATTATCCGCAGAA 300

  101 GlnAsnGluAsnLeuArgSerAlaPheSerAlaProThrSerAlaLeuPh 117
      |||||||||||||||||||||||||||||||||||||||||||||||||||
  301 CAAAATGAGAACCTGCGGAGCGCGTTTTCTGCGCCGACGTCGGCCTTATT 350

  117 eSerAlaSerProMetAlaGlnProArgThrThrIleSerAspAlaGluI 134
      |||||||||||||||||||||||||||||||||||||||||||||||||||
  351 TAGCGCTTCGCCTATGGCGCAGCCGAGAACAACCATTTCTGATGCTGAGA 400

  134 leTrpAspMetValSerGlnAsnIleSerAlaIleGlyAspSerTyrLeu 150
      |||||||||||||||||||||||||||||||||||||||||||||||||||
  401 TTTGGGATATGGTTTCCCAAAATATATCGGCGATAGGTGACAGCTATCTG 450

  151 GlyValTyrGluAsnValValAlaValTyrThrAspPheTyrGlnAlaPh 167
      |||||||||||||||||||||||||||||||||||||||||||||||||||
  451 GGCGTTTATGAAAACGTTGTCGCAGTCTATACCGATTTTTATCAGGCCTT 500

  167 eSerAspIleLeuSerLysMetGlyGlyTrpLeuLeuProGlyLysAspG 184
      |||||||||||||||||||||||||||||||||||||||||||||||||||
  501 CAGTGATATTCTTTCCAAAATGGGAGGCTGGTTATTACCAGGTAAGGACG 550

  184 lyAsnThrValLysLeuAspValThrSerLeuLysAsnAspLeuAsnSer 200
      |||||||||||||||||||||||||||||||||||||||||||||||||||
  551 GTAATACCGTTAAGCTAGATGTTACCTCACTCAAAAATGATTTAAACAGT 600

  201 LeuValAsnLysTyrAsnGlnIleAsnSerAsnThrValLeuPheProAl 217
      |||||||||||||||||||||||||||||||||||||||||||||||||||
  601 TTAGTCAATAAATATAATCAAATAAACAGTAATACCGTTTTATTTCCAGC 650

  217 aGlnSerGlySerGlyValLysValAlaThrGluAlaGluAlaArgGlnT 234
      |||||||||||||||||||||||||||||||||||||||||||||||||||
  651 GCAGTCAGGCAGCGGCGTTAAAGTAGCCACTGAAGCGGAAGCGAGACAGT 700

  234 rpLeuSerGluLeuAsnLeuProAsnSerCysLeuLysSerTyrGlySer 250
      |||||||||||||||||||||||||||||||||||||||||||||||||||
  701 GGCTCAGTGAATTGAATTTACCGAATAGCTGCCTGAAATCTTATGGATCC 750

  251 GlyTyrValValThrValAspLeuThrProLeuGlnLysMetValGlnAs 267
      |||||||||||||||||||||||||||||||||||||||||||||||||||
  751 GGTTATGTCGTCACCGTTGATCTGACGCCATTACAAAAAATGGTTCAGGA 800

  267 pIleAspGlyLeuGlyAlaProGlyLysAspSerLysLeuGluMetAspA 284
      |||||||||||||||||||||||||||||||||||||||||||||||||||
  801 TATTGATGGTTTAGGCGCGCCGGGAAAAGACTCAAAACTCGAAATGGATA 850

  284 snAlaLysTyrGlnAlaTrpGlnSerGlyPheLysAlaGlnGluGluAsn 300
      |||||||||||||||||||||||||||||||||||||||||||||||||||
  851 ACGCCAAATATCAAGCCTGGCAGTCGGGTTTTAAAGCGCAGGAAGAAAAT 900

  301 MetLysThrThrLeuGlnThrLeuThrGlnLysTyrSerAsnAlaAsnSe 317
      |||||||||||||||||||||||||||||||||||||||||||||||||||
  901 ATGAAAACCACATTACAGACGCTGACGCAAAAATATAGCAATGCCAATTC 950

  317 rLeuTyrAspAsnLeuValLysValLeuSerSerThrIleSerSerSerL 334
      |||||||||||||||||||||||||||||||:::|||||||||||||||||
  951 ATTGTACGACAACCTGGTAAAAGTGCTGAGCACTACGATAAGTAGCAGCC 1000

  334 euGluThrAlaLysSerPheLeuGlnGly 343
      |||||||||||||||||||||||||||||
 1001 TGGAAACCGCCAAAAGCTTCCTGCAAGGA 1029
```

```
seq_name: N_Geneseq_36:T67035

seq_documentation_block:
ID   T67035 standard; DNA; 3441 BP.
AC   T67035;
DT   04-AUG-1997  (first entry)
DE   Salmonella virulence factor SspBCDA DNA region.
KW   Salmonella secreted protein; SspB; SspC; SspD; SspA;
KW   bacterial-mediated endocytosis; diagnosis; therapy; vaccine;
KW   attenuation; virulence; ss.
OS   Salmonella typhimurium.
FH   Key             Location/Qualifiers
FT   cds             1. .870
FT                   /*tag=  a
FT                   /product= SspB (incomplete)
FT   cds             898. .2127
FT                   /*tag=  b
FT                   /product= SspC
FT   cds             2198. .3228
FT                   /*tag=  c
FT                   /product= SspD
FT   cds             3247. .3441
FT                   /*tag=  d
FT                   /product= SspA (incomplete)
PN   WO9718225-A1.
PD   22-MAY-1997.
PF   14-NOV-1996; U18504.
PR   14-NOV-1995; US-006733.
PA   (GEHO ) GEN HOSPITAL CORP.
PI   Miller SI;
DR   WPI; 97-289217/26.
DR   P-PSDB; W15268-71.
PT   New isolated Salmonella secreted proteins and related genes - used
PT   to develop products for the detection, treatment or prevention of
PT   Salmonella infections
PS   Claim 11; Fig 28; 95pp; English.
CC   A DNA molecule (T67035) comprises the SspBCDA locus (truncated at
CC   the 5' and 3' ends) including the partial sspB, sspC, sspD and
CC   partial sspA sequences (see also T67028-31) coding for Salmonella
CC   typhimurium secreted proteins SspB, SspC, SspD and SspA (W15268-
CC   70).  To identify the ssp genes, S. typhimurium mutants were
CC   investigated in invasion assays on HEp-2 epithelial cells.  2
CC   Ssp mutants with transposon insertions located between spaT and
CC   prgH (see also T67036) were identified.  One mutant lacked Ssp of
CC   87, 42 and 36 kDa.  The invasiveness of this mutant implicated the
CC   42 and 36 kDa Ssp in invasion.  DNA sequencing of this region
CC   identified the 4 ssp (partial) genes.  The ssp genes can be used in
CC   methods for diagnosing Salmonella infection and in methods of
CC   targetting an antigen to epithelial cells and of inducing a
CC   cytotoxic T cell immune response in a mammal.  Ssp gene mutation
CC   can be used to create attenuated Salmonella strains useful as live
CC   vaccines.
SQ   Sequence   3441 BP;    978 A;    761 C;    925 G;    777 T;


alignment_scores:
            Quality: 1445.00              Length:    343
              Ratio:  4.213                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 99.417

alignment_block:
 US-09-068-804-7 x T67035    ..

 Align seg 1/1  to: T67035  from: 1  to: 3441

       1 MetLeuAsnIleGlnAsnTyrSerAlaSerProHisProGlyIleValAl 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2198 ATGCTTAATATTCAAAATTATTCCGCTTCTCCTCATCCGGGGATCGTTGC 2247

      17 aGluArgProGlnThrProSerAlaSerGluHisValGluThrAlaValV 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2248 CGAACGGCCGCAGACTCCCTCGGCGAGCGAGCACGTCGAGACTGCCGTGG 2297
```

```
  34 alProSerThrThrGluHisArgGlyThrAspIleIleSerLeuSerGln 50
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2298 TACCGTCTACCACAGAACATCGCGGTACAGATATCATTTCATTATCGCAG 2347

  51 AlaAlaThrLysIleHisGlnAlaGlnGlnThrLeuGlnSerThrProPr 67
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2348 GCGGCTACTAAAATCCACCAGGCACAGCAGACGCTGCAGTCAACGCCACC 2397

  67 oIleSerGluGluAsnAsnAspGluArgThrLeuAlaArgGlnGlnLeuT 84
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2398 GATCTCTGAAGAGAATAATGACGAGCGCACGCTGGCGCGCCAGCAGTTGA 2447

  84 hrSerSerLeuAsnAlaLeuAlaLysSerGlyValSerLeuSerAlaGlu 100
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2448 CCAGCAGCCTGAATGCGCTGGCGAAGTCCGGCGTGTCATTATCCGCAGAA 2497

 101 GlnAsnGluAsnLeuArgSerAlaPheSerAlaProThrSerAlaLeuPh 117
     ||||||||||||||||||||||||||||||||||||||||||||||||
2498 CAAAATGAGAACCTGCGGAGCGCGTTTTCTGCGCCGACGTCGGCCTTATT 2546

 117 eSerAlaSerProMetAlaGlnProArgThrThrIleSerAspAlaGluI 134
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2547 .AGCGCTTCGCCTATGGCGCAGCCGAGAACAACCATTTCTGATGCTGAGA 2596

 134 leTrpAspMetValSerGlnAsnIleSerAlaIleGlyAspSerTyrLeu 150
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2597 TTTGGGATATGGTTTCCCAAAATATATCGGCGATAGGTGACAGCTATCTG 2646

 151 GlyValTyrGluAsnValValAlaValTyrThrAspPheTyrGlnAlaPh 167
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2647 GGCGTTTATGAAAACGTTGTCGCAGTCTATACCGATTTTTATCAGGCCTT 2696

 167 eSerAspIleLeuSerLysMetGlyGlyTrpLeuLeuProGlyLysAspG 184
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2697 CAGTGATATTCTTTCCAAAATGGGAGGCTGGTTATTACCAGGTAAGGACG 2746

 184 lyAsnThrValLysLeuAspValThrSerLeuLysAsnAspLeuAsnSer 200
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2747 GTAATACCGTTAAGCTAGATGTTACCTCACTCAAAAATGATTTAAACAGT 2796

 201 LeuValAsnLysTyrAsnGlnIleAsnSerAsnThrValLeuPheProAl 217
     ||||||||||||||||||||||||||||||:::||||||||||||||||
2797 TTAGTCAATAAATATAATCAAATAAACAGTAATAGCGTTTTATTTCCAGC 2846

 217 aGlnSerGlySerGlyValLysValAlaThrGluAlaGluAlaArgGlnT 234
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2847 GCAGTCAGGCAGCGGCGTTAAAGTAGCCACTGAAGCGGAAGCGAGACAGT 2896

 234 rpLeuSerGluLeuAsnLeuProAsnSerCysLeuLysSerTyrGlySer 250
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2897 GGCTCAGTGAATTGAATTTACCGAATAGCTGCCTGAAATCTTATGGATCC 2946

 251 GlyTyrValValThrValAspLeuThrProLeuGlnLysMetValGlnAs 267
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2947 GGTTATGTCGTCACCGTTGATCTGACGCCATTACAAAAAATGGTTCAGGA 2996

 267 pIleAspGlyLeuGlyAlaProGlyLysAspSerLysLeuGluMetAspA 284
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2997 TATTGATGGTTTAGGCGCGCCGGGAAAAGACTCAAAACTCGAAATGGATA 3046

 284 snAlaLysTyrGlnAlaTrpGlnSerGlyPheLysAlaGlnGluGluAsn 300
     ||||||||||||||||||||||||||||||||||||||||||||||||||
3047 ACGCCAAATATCAAGCCTGGCAGTCGGGTTTTAAAGCGCAGGAAGAAAAT 3096

 301 MetLysThrThrLeuGlnThrLeuThrGlnLysTyrSerAsnAlaAsnSe 317
     ||||||||||||||||||||||||||||||||||||||||||||||||||
3097 ATGAAAACCACATTACAGACGCTGACGCAAAAATATAGCAATGCCAATTC 3146

 317 rLeuTyrAspAsnLeuValLysValLeuSerSerThrIleSerSerSerL 334
     ||||||||||||||||||||||||||||||||||||||||||||||||||
3147 ATTGTACGACAACCTGGTAAAAGTGCTGAGCAGTACGATAAGTAGCAGCC 3196

     334 euGluThrAlaLysSerPheLeuGlnGly 343
         |||||||||||||||||||||||||||
    3197 TGGAAACCGCCAAAAGCTTCCTGCAAGGA 3225
```

```
seq_name: N_Geneseq_36:X39703

seq_documentation_block:
ID   X39703 standard; DNA; 744 BP.
AC   X39703;
DT   02-JUL-1999  (first entry)
DE   Gastric cancer associated gene.
KW   Cancer associated antigen; diagnosis; research; treatment; human;
KW   breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW   prostate cancer; ss.
OS   Homo sapiens.
PN   WO9904265-A2.
PD   28-JAN-1999.
PF   15-JUL-1998; U14679.
PR   22-JUN-1998; US-102322.
PR   17-JUL-1997; US-896164.
PR   10-OCT-1997; US-061599.
PR   10-OCT-1997; US-061765.
PR   10-OCT-1997; US-948705.
PR   11-OCT-1997; GB-021697.
PA   (LUDW-) LUDWIG INST CANCER RES.
PI   Chen Y, Gout I, Gure A, OHare M, Obata Y, Old LJ,
PI   Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E,
PI   Tureci O;
DR   WPI; 99-132448/11.
PT   New isolated cancer associated nucleic acids and polypeptides -
PT   isolated using sera from cancer patients, used to develop products
PT   for the diagnosis, monitoring or treatment of cancers
PS   Claim 67; Page 516-517; 787pp; English.
CC   The invention relates to a method for diagnosing a disorder characterised
CC   by expression of a human cancer associated antigen precursor coded for by
CC   a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC   biological sample isolated from a subject with an agent that specifically
CC   binds to the NAM, an expression product or a fragment of an expression
CC   product complexed with an HLA molecule; and (b) determining the
CC   interaction between the agent and the NAM or the expression product as a
CC   determination of the disorder. The products and methods can be used in
CC   the diagnosis, monitoring, research, or treatment of conditions
CC   characterised by the expression of various cancer associated antigens.
CC   The invention provides nucleic acid sequences and encoded polypeptides
CC   which are cancer associated antigen precursors expressed in human breast
CC   cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC   lung cancer.
SQ   Sequence   744 BP;    194 A;    198 C;    155 G;    197 T;


alignment_scores:
           Quality:   77.50              Length:    141
             Ratio:  0.861                Gaps:      4
 Percent Similarity:  63.830   Percent Identity: 17.730

alignment_block:
 US-09-068-804-7 x X39703   ..

 Align seg 1/1  to: X39703  from: 1  to: 744

       8 SerAlaSerProHisProGlyIleValAlaGluArgProGlnThrProSe 24
         ::::::::::::::::   :::   ::::::::::::::::::::::|||::
      72 CCTACTGGCAGTCAGATTGCAAATATTGGTCAGCAAGCAAACATACCTAC 121

      24 rAlaSerGluHisValGluThrAlaValValProSerThrThrGluHis. 40
         :|||   ::::::      |||   |||   ||||||:::::::::::::
     122 TGCAGTGCAGCAGCCCTCTACCCAGGTTCCACCTTCAGTTATTCAGCAGG 171

      41 ........ArgGlyThrAspIleIleSerLeuSerGlnAlaAlaThrLys 54
                 ::::::::::::::::::   :::|||:::::::::
     172 GTGCTCCTCCATCTTCGCAAGTGGTTCCACCTGCTCAAACTGGGATT... 218

      55 IleHisGlnAlaGlnGlnThr.......................... 61
```

```
      |||||||||:::   ||||||
    219 ATTCATCAGGGAGTTCAAACTAGTGCTCCAAGCCTTCCTCAACAATTGGT 268

     62 ..................LeuGlnSerThrProProIleSerGluGlu. 71
                           |||   ::::::||||||   ::::::
    269 TATTGCATCCCAAAGTTCCTTGTTAACTGTGCCTCCCCAGCCACAAGGAG 318

     72 ..AsnAsnAspGluArgThrLeuAlaArgGlnGlnLeuThrSerSerLeu 87
          ::::::   ::::::   ::::::   |||||||||:::::::
    319 TAGAATCAGTAGCTCAAGGAATTGTTTCACAGCAGTTGCCTGCAGTTAGT 368

     88 AsnAlaLeuAlaLysSerGlyValSerLeuSerAlaGluGlnAsnGluAs 104
        :::      :::   |||:::::::|||:::::::::::::   :::   ::
    369 TCTTTGCCCTCTGCTAGTAGTATTTCTGTTACAAGTCAGGTTAGTTCAAC 418

    104 nLeuArgSerAlaPheSerAlaProThrSerAlaLeuPheSerAlaSerP 121
        :      |||:::   ::::::::::::::::::|||   ::::::
    419 TGGTCCTTCTGGAATGCCTTCTGCCCCAACAAACTTGGTTCCACCACAAA 468

    121 roMetAlaGlnProArgThrThr 128
          :::||||||:::   :::|||
    469 ATATAGCACAAACCCCTGCTACC 491

seq_name: N_Geneseq_36:X20263

seq_documentation_block:
ID   X20263 standard; DNA; 8905 BP.
AC   X20263;
DT   04-MAY-1999  (first entry)
DE   Borrelia burgdorferi polynucleotide sequence #16.
KW   Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
KW   epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
KW   infection; diagnosis; characterisation; detection; ds.
OS   Borrelia burgdorferi.
PN   WO9858943-A1.
PD   30-DEC-1998.
PF   18-JUN-1998; U12764.
PR   03-SEP-1997; US-057483.
PR   20-JUN-1997; US-050359.
PR   22-JUL-1997; US-053344.
PR   22-JUL-1997; US-053377.
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (MEDI-) MEDIMMUNE INC.
PI   Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO,
PI   White OR;
DR   WPI; 99-081217/07.
PT   New isolated Borrelia burgdorferi nucleic acids - used to develop
PT   products for the detection, diagnosis, characterisation, prevention
PT   and therapy of infections, particularly Lyme disease
PS   Claim 1; Page 936-941; 1128pp; English.
CC   X20248 to X20402 represent polynucleotide sequences isolated from
CC   Borrelia burgdorferi (Bb). Products derived from Bb can be used for
CC   the detection, diagnosis, characterisation, prevention and therapy of
CC   Bb infections, e.g. Lyme disease. They can also be used for the
CC   production of biosynthetic products, e.g. enzymes. Borrelia belongs
CC   to a family of motile, spiral-shaped bacteria called Spirochetes.
CC   Spirochetes are pathogenic in humans and Borrelia causes epidemic and
CC   endemic relapsing fever, and Lyme borreliosis, more commonly known as
CC   Lyme disease.
SQ   Sequence   8905 BP;    3157 A;    1108 C;    1183 G;    3455 T;


alignment_scores:
            Quality:   85.50            Length:    294
              Ratio:  0.590              Gaps:     10
 Percent Similarity:  49.320   Percent Identity: 18.707

alignment_block:
 US-09-068-804-7 x X20263/rev   ..

 Align seg 1/1  to reverse of: X20263  from: 1  to: 8905

     23 ProSerAlaSerGluHisValGluThrAlaValValProSerThrThrGl 39
        :
      |||::::::      ::::::|||:::||||||::|||
   4903 CCCACTACTGTAGGCAATATAGAGTCCGCAGTCCTCCCTGTA..... 4862

     39 uHisArgGlyThrAspIleIleSerLeuSerGlnAlaAlaThrLysIleH 56
                  ::::::::::|||||      ::::|||:::|||:::
   4861 ..........GGGCATGTAGTGTCTTTAGAGACATCTGCTAATAAGGTT. 4823

     56 isGlnAlaGlnGlnThrLeuGlnSerThrProProIleSerGluGluAsn 72
                              |||:::|||:::||||||      :::
   4822 ......................AGTATACCAACCATTTCAATTAAGCAC 4796

     73 AsnAspGluArgThrLeuAlaArgGlnGlnLeuThrSerSerLeuAsnAl 89
        |||:::   :::   :::   ::::::::::|||:::::::|||   :::::
   4795 AATCAAAAAAAAAGAGATAAAAAAAGAAGATCTTTCCCCTTCTACTAAGGA 4746

     89 aLeuAlaLysSerGlyValSerLeuSerAlaGluGlnAsnGluAsnLeuA 106
        :      |||::::::   ::::::   :::   :::|||
   4745 AGAAAAGAAAGCAGATAAAGCAATTAAAGATATAGAGAATCTTATTAGAG 4696

    106 rgSerAlaPheSerAlaProThrSerAlaLeuPheSerAlaSerProMet 122
          |||:::|||::::::   :::   ::::::::::|||          :::
   4695 ACTCTGGATTTCCCGAGTTAATTGAGAGTATGTATTCA.....CTT 4655

    123 AlaGlnProArgThrThrIleSerAspAlaGluIleTrpAspMetValSe 139
        :::      |||   |||   :::
   4654 AAACATGAATATACTTTAATAAGAAAT..................... 4628

    139 rGlnAsnIleSerAlaIleGlyAspSerTyrLeuGlyValTyrGluAsnV 156

   4628 ..................................................... 4628

    156 alValAlaValTyrThrAspPheTyrGlnAlaPheSerAspIleLeuSer 172
                      :::||||||:::::::::::   |||   :::
   4627 ................AATTTCTATGATGTGATAACTAAAATTAGGAAC 4595

    173 LysMetGlyGlyTrpLeuLeuProGlyLysAspGlyAsnThrValLysLe 189
        |||:::   :::   :::      |||:::::::::
   4594 AAAAAAACATCATTAATAAAAAATGGTCGTAACAATAGAGATAAAATAAA 4545

    189 uAspValThrSerLeuLysAsnAspLeuAsnSerLeuValAsnLysTyrA 206
         ::::::|||   |||:::|||:::|||:::   :::|||:::      :
   4544 GGAACTAACACAATTGCAAAATAATTTAAAG...ATAGTGGATGAACTTG 4498

    206 snGlnIleAsnSerAsnThrValLeuProAlaGlnSerGlySerGly 222
        ::::::|||      :::
   4497 ATGAAATTATGGTTCAC..................................... 4481

    223 ValLysValAlaThrGlu..........................AlaGl 230
        :::   :::|||   :::                          :::||
   4480 ATTGATATTGCAGAACAAGAGATAAGATCTGCGGCTTTCTTTTTTAATGA 4431

    230 uAlaArgGlnTrpLeuSerGluLeuAsnLeuProAsnSerCysLeuLysS 247
        ||||::::::   |||   |||                :::   :::|||
   4430 AGCTAAGGAAATTTTAAAAGAA..............GGCATTATTAAAA 4396

    247 erTyrGlySerGlyTyrValValThrValAspLeuThr......ProLeu 261
              |||         |||:::   :::|||:::      :::|||
   4395 GATTGGAAAGTGAAAATAAAGTGGCATCACAATTAGCTAGACAGGCTTTA 4346

    262 GlnLysMetValGlnAspIleAspGlyLeuGlyAlaProGlyLysAspSe 278
        :::|||:::   :::::::::   :::|||   |||::::::
   4345 AATAAAGTAGAAGATGCTTTAAAGAGCTTAGAAGCTTCTTCTTCTAAAAG 4296

    278 rLysLeuGluMetAspAsnAlaLysTyrGlnAlaTrpGlnSerGlyPheL 295
           |||:::|||                 :::::::::      :::|
   4295 AGGTTTGGCCATG.................GGAAGAAGGAGAATTATAA 4264

    295 ysAlaGlnGluGluAsnMetLysThrThrLeu 305
        |||:::      ||||||   ||||||:::|||
   4263 AAGAACTTATTGAAAATGCAAAGACTGTTTTA 4232

seq_name: N_Geneseq_36:T14053
```

```
seq_documentation_block:
ID   T14053 standard; DNA; 2899 BP.
AC   T14053;
DT   24-JUL-1997  (revised)
DT   02-NOV-1996  (first entry)
DE   Fungal signal recognition particle SRP54 gene genomic clone.
KW   Signal recognition particle; recombinant; genetic engineering;
KW   secretion; filamentous fungi; transformation; Aspergillus;
KW   fungal enzyme; ds.
OS   Aspergillus niger.
FH   Key             Location/Qualifiers
FT   5'utr           1. .147
FT                   /*tag= a
FT   exon            148. .225
FT                   /*tag= b
FT                   /number= 1
FT   intron          226. .274
FT                   /*tag= c
FT                   /number= 1
FT   exon            275. .1801
FT                   /*tag= d
FT                   /number= 2
FT                   /note= "includes stop codon"
FT   3'utr           1802. .2899
FT                   /*tag= e
PN   WO9610635-A.
PD   11-APR-1996.
PF   02-OCT-1995; U12628.
PR   03-OCT-1994; US-317401.
PA   (NOVO ) NOVO NORDISK BIOTECH INC.
PI   Thompson SA,  Yaver DS;
DR   WPI; 96-209357/21.
DR   P-PSDB; R91309.
PT   Nucleic acid construct encoding elements of signal recognition particle
PT   of filamentous fungi - used to increase prodn. of heterologous proteins
PT   in fungi.
PS   Claim 7; Fig 2; 39pp; English.
CC   T14053 and T14054 both encode an Aspergillus niger signal recognition
CC   particle element, SRP54, derived from a genomic Bo1 clone and a cDNA
CC   SFAG2 clone respectively. A nucleic acid construct contg. DNA encoding
CC   SRP54 may be used for the recombinant production of a protein of
CC   interest, in particular for the production of fungal enzymes e.g.
CC   catalase, laccase, cellulase, xylanase, glucoamylase, pectinase or
CC   galactosidase. The SRP54 DNA increases the level of secretory
CC   components in transformed cells, improving the production of secreted
CC   proteins.
CC   (Revised entry submitted to correct sequence analysis breakdown and
CC   error in the Feature Table.)
SQ   Sequence   2899 BP;    634 A;    805 C;    706 G;    685 T;


alignment_scores:
            Quality:   79.50              Length:    346
              Ratio:  0.437                Gaps:     15
 Percent Similarity:  52.601   Percent Identity: 19.075

alignment_block:
 US-09-068-804-7 x T14053   ..

 Align seg 1/1  to: T14053  from: 1  to: 2899

       3 AsnIleGlnAsnTyrSerAlaSerProHis......ProGlyIleValAl 17
         :::|||:::   |||:::::::   ::::::      |||:::|||||||||
     665 AAGATCCCCTACTACGGTAGCCTGACGCAAACCGACCCCGCCATCGTGGC 714

      17 aGluArgProGlnThrProSerAlaSerGluHisValGluThrAlaValV 34
         |:::       :::            |||:::   |||:::   :::|
     715 AGCCGAGGGTGTGGCCAAGTTCAAGAAGGAGCGTTTCGAAATCATCATTG 764

      34 alProSerThrThrGluHisArgGlyThrAsp......IleIleSerLeu 48
         ||   ::::::      |||:::      :::      ::::::   :::
     765 TCGATACCAGTGGTCGTCATAAGCAGGAAGAAGAGCTCTTCACCGAAATG 814

      49 SerGlnAlaAlaThrLysIleHisGlnAlaGlnGlnThrLeuGlnSerTh 65
         :::|||      |||   :::   ::::::|||   :::|||
     815 ACCCAGATTCAGACCGCCGTCACCCCCGACCAGACCATCCTCGTCCTCGA 864

      65 rProProIleSerGluGluAsnAsnAspGluArgThrLeu.AlaArgGln 81
         ::::::|||:::::::::::::::::::::::   :::    :::|||:::
     865 CAGCACCATCGGTCAGGCTGCCGAAGCCCAGTCCTCCGCCTTCAAGGCCA 914

      82 GlnLeuThrSerSerLeuAsnAlaLeuAlaLysSerGlyValSerLeuSe 98
         :::   ||||||
     915 CCGCAGACTTCG..................................... 926

      98 rAlaGluGlnAsnGluAsnLeuArgSerAlaPheSerAlaProThrSerA 115
            |||::::::   :::   |||         ::::::::::::   :
     927 ....GAGCCATCATCATCACCAAGACGGATGGTCACGCCGCAGGTGGTGG 972

     115 laLeuPheSerAlaSerPro............................. 121
         ::|||||||:::::|||||||
     973 TGCTATTTCCGCCGTCGCCGCCACACACACTCCCATTATCTACCTCGGTA 1022

     122 .........................MetAlaGlnProArgThrThrIl 129
                                  ::::::   ||||||::::::
    1023 CCGGTGAGCACTTGATGGACCTGGAACGTTTCGAGCCCAAGGCCTTCATC 1072

     129 eSerAspAlaGluIleTrp..AspMetValSer..............Gl 140
         ::::::   :::|||  ||||||::::::                ||
    1073 CAGAAGCTCCTGGGTATGGGTGATATGGCCGGACTGGTAGAGCACGTACA 1122

     140 nAsnIleSerAlaIleGlyAspSerTyrLeuGlyValTyrGluAsnValV 157
         |:::::::::      :::::::|||          :::|||   ::::::
    1123 AGCCGTGACCAAGGACTCGGCCTCCGCCAAGGAAACATACAAGCACATCT 1172

     157 al...AlaValTyrThr......AspPheTyrGlnAlaPheSerAspIle 170
             ::::::||||||      ||||||   ::::::::::::::|||
    1173 CAGAAGGTATCTACACGCTGCGCGACTTCCGCGAGAACATCACCTCCATC 1222

     171 ..............LeuSerLysMetGlyGlyTrpLeuLeuProGlyLy 182
                       |||||||||::::::|||   ::::::||||||
    1223 ATGAAGATGGGTCCTCTCTCCAAGCTCTCCGGC...ATGATTCCCGGTCT 1269

     182 sAspGlyAsnThrValLysLeuAspValThrSerLeuLysAsnAspLeuA 199
         ::::::   |||:::   ||||||      :::             |||
    1270 CTCCAACCTGACCGCGGGACTTGATGACGAAGACGGCTCCATGAAGCTCC 1319

     199 snSerLeuValAsnLysTyrAsn..............GlnIleAsnSer 210
             ::::::      ::::::              ::::::::::::
    1320 GCCGCATGATCTACATTTTCGACAGCATGACGGCCGCCGAACTCGACGGC 1369

     211 AsnThrValLeuPheProAlaGlnSerGlySerGlyValLysValAla.. 226
         :::       :::|||   :::|||:::::::      |||::::::|||
    1370 GACGGCAAGATGTTCGTCGAACAGCCTAGCCGCATGGTCCGGATCGCTTG 1419

     227 .............ThrGluAlaGluAlaArgGlnTrpLeuSerGluLeuA 239
                       |||      |||:::   :::   |||||:::
    1420 CGGAAGCGGTACCACCGTCCGCGAAGTCGAAGACCTGCTCTCCCAGCACC 1469

     239 snLeuProAsnSerCysLeuLysSerTyrGlySer.............. 250
           :::   ::::::      |||      |||:::
    1470 GCATGATGGCTGGTATGGCCAAGCGCGTCGGCGGACAGAAGAAGCAGATG 1519

     251 .....................GlyTyrValValThrValAspLeuTh 259
                              |||               :::|||::
    1520 CAGCGTGCCCAGAACATGCTCAAGGGCGGTAACAAGGAACAGCAGCTCGC 1569

     259 rProLeuGlnLysMetValGlnAspIleAspGlyLeuGlyAla..... 273
         :::::::||||||   :::   :::::::::|||   |||:::
    1570 CGCCATGCAGAAGCGCATGGCCGCCATGGGTGGTGCTGGCGGTGGTGGAT 1619

     274 ..ProGlyLysAspSerLysLeuGluMetAspAsn 284
         ||||||:::   ::::::   :::|||::::::
    1620 TCCCCGGTATGCCCGGCATGGGCGACATGGCCAAG 1654

seq_name: N_Geneseq_36:V60373
```

```
seq_documentation_block:
ID   V60373 standard; cDNA; 1374 BP.
AC   V60373;
DT   16-DEC-1998  (first entry)
DE   Human Dil2 protein coding sequence.
KW   Dil2; breast cancer; cancer diagnosis; infiltrative ductal carcinoma;
KW   gene therapy; human; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             149. .1168
FT                   /*tag= a
FT                   /transl_except= (pos: 560. .562, aa: Asn)
FT                   /product= "Dil2"
PN   WO9842725-A1.
PD   01-OCT-1998.
PF   20-MAR-1998; U05629.
PR   21-MAR-1997; US-044425.
PA   (MUSC-) MUSC FOUND RES DEV.
PI   Papas TS, Seth AK, Vournakis JN;
DR   WPI; 98-531945/45.
DR   P-PSDB; W71427.
PT   New nucleic acid encoding the breast cancer associated protein Dil2
PT   - and related vectors, transformed cells, antibodies, proteins and
PT   inhibitors, useful for diagnosis, staging, monitoring and treating
PT   breast cancer
PS   Claim 3; Fig 1B; 105pp; English.
CC   This sequence represents the DNA of the invention, and encodes the human
CC   Dil2 protein. Dil2 is expressed in breast cancer tissues and cell lines.
CC   The Dil2 protein is detected in breast cancer, but not in benign disease
CC   or normal tissue, and is markedly increased where metastasis or lymph
CC   node involvement is present. Cells transformed with the DNA are used to
CC   produce recombinant Dil2. Fragments of the DNA are useful as probes and
CC   primers, in standard hybridisation and amplification assays, to detect,
CC   quantify or isolate the DNA and its mutants, for diagnosis of breast
CC   cancer or identification of subjects at risk. Antibodies against the Dil2
CC   protein are used as immunoassay reagents to detect the Dil2 gene product,
CC   for diagnosis of breast cancer (particularly infiltrative ductal
CC   carcinoma), also for staging (i.e. detecting involvement of lymph nodes
CC   or metastases) and monitoring response to therapy. Agents that modulate
CC   Dil2 gene expression, that decrease expression of Dil2 or antagonise its
CC   activity, also Dil2 antisense, ribozyme or triplex-forming molecules, are
CC   useful for treatment of breast cancer (including expression in gene
CC   therapy systems). The protein can be used to raise antibodies and to
CC   identify other cellular gene products involved in development of cancer.
SQ   Sequence   1374 BP;    408 A;    270 C;    330 G;    366 T;


alignment_scores:
            Quality:   75.00              Length:    330
              Ratio:  0.399                Gaps:     14
 Percent Similarity:  56.970   Percent Identity: 17.273

alignment_block:
 US-09-068-804-7 x V60373   ..

 Align seq 1/1  to: V60373  from: 1  to: 1374

      30 GluThrAlaValValProSerThrThrGluHisArgGlyThrAspIleIl 46
         ::::::||||||:::::::::   |||         ::::::||||||
      29 CAAGCCGCAGTCATCGCTGGTGGTACA.........TCAAAAGACATTAA 69

      46 eSerLeuSerGlnAlaAlaThrLysIleHisGlnAlaGlnGlnThrLeuG 63
          :::|||:::::::::|||:::|||:::
      70 CGGACTTGCTGACGTAGCCAACAAAATG..................... 97

      63 lnSerThrProProIleSerGluGluAsnAsnAspGluArgThrLeuAla 79
                                  :::::::::   :::|||:::
      98 ............................GGTGCTGACCTGCCTTTAAGC 118

      80 ArgGlnGlnLeuThrSerSerLeuAsnAlaLeuAlaLysSerGlyValSe 96
                    :::   ::::::   |||:::|||:::::::|||:::||
     119 GCTGGTGCTGCTGCGAAAGCTATGGTAGCCATGGCTCGTGACGGAGCGTC 168

      96 rLeuSerAlaGluGlnAsnGluAsnLeuArgSerAlaPheSerAlaProT 113
         |:::   :::                 :::::::|||:::|||:::|||   :
     169 TATCAAGACA..............ATTAAATCAGAGTTCCCAGCTATTG 203

     113 hrSerAlaLeuPheSerAlaSerProMetAlaGlnProArgThrThrIle 129
         ::   |||      :::|||::::::      |||:::   :::::::::
     204 CGCAAGCTGCAACGGCTGCTGGTGCTGATTTGCAGCAAACAGCGTCAGTT 253

     130 SerAsp......AlaGluIleTrpAspMetValSerGlnAsnIleSerAl 144
            :::         :::|||||:::         |||:::
     254 GTTCAACAATCGATGAACATCTGGGGTGATAGTATTCAAAGC........ 295

     144 aIleGlyAspSerTyrLeuGlyValTyrGluAsnValValAlaValTyrT 161
               :::         :::|||      ::::::::::::::   :
     296 ....CCACAACGTGCTGCTGCTGTATTAACGCAAACTGCTAACCTATCCA 341

     161 hrAspPheTyrGlnAlaPheSerAspIleLeuSerLysMetGlyGlyTrp 177
         :::::      ::::::      :::   |||:::::::::||||||
     342 ACGCCTCAATCGAAGATATGCAACAGGCTTTAGCTACTATCGGTGGT... 388

     178 LeuLeuProGlyLysAspGlyAsnThrValLysLeuAspValThrSerLe 194
                              ::::::::::::   :::|||:::   :::
     389 ..............ACGGCACACAACGCTGGTATCGACATGCAAACAAC 423

     194 uLysAsnAspLeuAsnSerLeuValAsnLys..................T 205
            ::::::::::::   |||:::|||:::
     424 GTCTACCGCTATCGGTTTACTTACTAACCGTGGTTTCAGTGCTGCACAAG 473

     205 yrAsnGlnIleAsnSerAsnThrValLeuPhe.........ProAlaGln 218
            :::|||      :::::::::::|||:::         |||::::::
     474 CGTCACAAGACTTGAACCATGCTTTGTTGCTCATGCAAGCTCCAAGTGAA 523

     219 SerGlySerGlyValLysValAlaThrGluAlaGluAlaArgGlnTrpLe 235
            |||:::|||||||                                ::
     524 AAAGGAGCTGGTGTG................................AT 540

     235 uSerGluLeuAsnLeuProAsnSerCysLeuLysSerTyrGlySerGlyT 252
         :   :::|||:::|||:::   :::
     541 GCACAATCTCGGTTTGTCTATGACA........................... 565

     252 yrValValThrValAspLeuThrProLeuGlnLysMetValGlnAspIle 268
            :::      :::::::::||||||:::|||:::::::::::::|||
     566 ..GACGCACAAGGAAACATGAAGCCGTTACCAAAAATTTTGAATGAAATT 613

     269 Asp............GlyLeuGlyAlaProGlyLysAspSerLysLeuGl 281
         :::             |||:::   ::::::::::|||::::::   |||
     614 GGTGACGCGACACGTGGGATGACTAGTTCTGATAAAGCAGCAGCGTTG.. 661

     281 uMetAspAsnAlaLysTyrGlnAlaTrpGlnSerGlyPheLysAlaGlnG 298
               :::|||:::::   :::                    |||
     662 .......AAGGCTATGTTTGGTACT..................GCTGGTA 686

     298 luGluAsnMetLysThrThrLeuGlnThrLeuThrGlnLysTyrSerAsn 314
            :::::::::   :::   ::::::::::::::::::|||   :::|||
     687 TGGCTGCTATTCTTCCTTTGATGGATAGTGTCAAGGATAAGACTGGTAAT 736

     315 AlaAsnSerLeuTyrAspAsnLeuValLysValLeuSerSerThrIleSe 331
         |||::::::      :::::::::::|||   ::::::   :::   ||
     737 GCTACCACAAGTTGGGAAGCTTTCACCAAAGAAATGGATAAAGCG...TC 783

     331 rSerSerLeuGluThrAlaLysSerPheLeuGlnGly*** 344
         |:::|||   :::||||||:::::::||||||:::::|||
     784 AGGAAGTACGCAAACAGCTACAAACTTCCTAAAAGATCAA 823

seq_name: N_Geneseq_36:X09011

seq_documentation_block:
ID   X09011 standard; cDNA; 1266 BP.
AC   X09011;
DT   14-JUN-1999  (first entry)
DE   Brn-3a polynucleotide.
KW   Brn-3a; Bcl-2; neurons; neuronal cells; apoptosis; cell death; CNS;
```

```
KW   PNS; central nervous system; parasympathetic nervous system;
KW   development; injury; neurotrophic factor; nerve growth factor; NGF;
KW   ciliary neurotrophic factor; CNTF; brain-derived neurotrophic factor;
KW   BNTF; neurotrophin; NT-3; NT-4; NT-5; neurodegenerative disease;
KW   familial dysautonomia; infantile muscular dystrophy;
KW   Parkinson's disease; Alzheimer's disease; ss.
OS   Mus musculus.
FH   Key             Location/Qualifiers
FT   CDS             1. .1266
FT                   /*tag= a
FT                   /product= Brn_3a_polypeptide
PN   WO9905272-A1.
PD   04-FEB-1999.
PF   27-JUL-1998; G02228.
PR   10-DEC-1997; US-988476.
PR   25-JUL-1997; GB-015823.
PA   (UNLO ) UNIV COLLEGE LONDON.
PI   Latchman DS, Smith MD;
DR   WPI; 99-142928/12.
DR   P-PSDB; W96263.
PT   New polypeptide comprising transcription factor Brn-3a, or its
PT   derivative - useful for treating nervous system diseases, preventing
PT   cellular apoptosis and increasing nerve regeneration following
PT   neuronal damage
PS   Disclosure; Page 63-64; 68pp; English.
CC   Over expression of transcription factor Brn-3a can protect neuronal
CC   cells from apoptosis. Brn-3a also specifically activates expression
CC   of the BCl-2 gene in neuronal cells and this activation is mediated
CC   via a Brn-3a response element in the 5' regulatory region of the
CC   Bcl-2 gene. Both the anti-apoptopic effect of Brn-3a and its ability
CC   to activate expression of Bcl-2 are mediated by the N-terminal
CC   domain of Brn-3a. Members of the Bcl-2 family perform critical roles
CC   in the regulation of selective apoptosis during development of the
CC   nervous system.  The stimulation of Bcl-2 expression by Brn-3a in a
CC   neuron specific manner and consequent protection of neuronal cells
CC   from apoptosis suggests that Brn-3a may co-ordinate some aspects of
CC   neuronal reorganisation during development or following injury.  The
CC   elevation of Brn-3a expression by either pharmacological means
CC   (compositions comprising one other therapeutic polypeptide e.g.
CC   neurotrophic factors, nerve growth factor (NGF), ciliary neurotrophic
CC   factor (CNTF), brain-derived neurotrophic factor (BNTF), and
CC   neurotrophins NT-3 and NT-4/5) or gene therapy may represent a
CC   method for treating human diseases associated with excessive
CC   neuronal cell death and/or lack of nerve regeneration, especially
CC   neurodegenerative diseases such as familial dysautonomia and
CC   infantile muscular dystrophy, and Parkinson's and Alzheimer's
CC   disease.
SQ   Sequence   1266 BP;    201 A;    455 C;    458 G;    152 T;


alignment_scores:
            Quality:   74.50              Length:    189
              Ratio:  0.665                Gaps:      8
 Percent Similarity:  59.259    Percent Identity: 20.106

alignment_block:
 US-09-068-804-7 x X09011   ..

 Align seg 1/1  to: X09011  from: 1  to: 1266

      11 ProHisProGlyIleValAlaGluArgProGln..................  21
         ||||||||||||:::|||||||:::   ::::::
     649 CCGCATCCCGGGCTCGTGGCCGCGGCGGCGCACCACGGCGCGGCGGCGGC 698

      22 ....ThrProSerAlaSerGluHisValGluThrAlaValValProSerT  37
             ::::::::::|||:::   :::|||::::::|||   ::::::::::
     699 AGCGGCGGCGGCGGCGGCGGGGCAGGTGGCGGCGGCGTCGGCCGCGGCGG 748

      37 hrThr..GluHisArgGlyThrAspIle...IleSerLeuSerGlnAlaA  52
         :::::     :::   |||:::::::::   :::   |||::::::::::
     749 CGGTGGTGGGCGCGGCGGGCCTGGCGTCCATCTGCGACTCGGACACGGAC 798

      52 laThrLysIleHisGlnAlaGlnGlnThrLeuGlnSerThrProProIle  68
         ::::::::   :::   ::::::::::::|||||:::::::::
     799 CCGCGCGAGCTCGAGGCGTTCGCCGAGCGCTTCAAGCAGCGGCGCATCAA 848

      69 SerGluGluAsnAsnAspGluArgThrLeuAlaArgGlnGlnLeuThrSe  85
         :::      :::::::::   |||   |||:::      |||      ::
     849 GCTGGGCGTGACGCAGGCCGACGTGGGCTCGGCGCTGGCCAACCTCAAGA 898

      85 rSerLeuAsnAlaLeuAlaLysSerGlyValSerLeuSer..........  98
         ::::      :::||||||:::::       |||
     899 TCCCGGGCGTGGGCTCGCTCAGCCAGAGCACCATCTGCAGGTTCGAGTCG 948

      99 ...........AlaGluGlnAsnGlu.AsnLeuArgSerAlaPheSerAl 111
                    |||:::|||:::::: :::|||:::::::   :::   ||
     949 CTCACGCTCTCGCACAACAACATGATCGCGCTCAAGCCCATCCTGCAGGC 998

     111 a.....................................ProThrSerA 115
         |                                     :::::::
     999 GTGGCTGGAGGAGGCCGAGGGCGCGCAGCGTGAGAAAATGAACAAGCCGG 1048

     115 laLeuPheSerAlaSerProMetAlaGlnProArgThrThrIleSerAsp 131
         ::|||||:::::::::   :::   :::   ||||||:::|||:::::
    1049 AGCTCTTCAACGGCGGCGAGAAGAAGCGCAAGCGGACTTCCATCGCCGCG 1098

     132 AlaGluIleTrpAspMet..............................Va 138
         :::|||   :::::::::
    1099 CCCGAGAAGCGCTCCCTCGAGGCCTATTTTGCCGTACAACCCCGGCCCTC 1148

     138 lSerGlnAsnIleSerAlaIleGlyAspSerTyrLeuGlyValTyrGluA 155
          |||:::::::|||:::||||||:::::      |||::::::      |
    1149 GTCTGAGAAGATCGCCGCCATCGCCGAGAAA...CTGGACCTCAAAAAGA 1195

     155 snValValAlaVal 159
         ||||||||   |||
    1196 ACGTGGTGCGGGTG 1209

seq_name: N_Geneseq_36:T68058

seq_documentation_block:
ID   T68058 standard; DNA; 1215 BP.
AC   T68058;
DT   16-JUL-1997  (first entry)
DE   H. pylori cytoplasmic protein ORF 09ap20802orf22.
KW   Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW   identification; binding compound; bacterium; life cycle; activator;
KW   bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;
KW   diagnosis; ds.
OS   Helicobacter pylori.
FH   Key             Location/Qualifiers
FT   cds             1. .1215
FT                   /*tag= a
FT                   /note= "no stop codon given"
PN   WO9640893-A1.
PD   19-DEC-1996.
PF   06-JUN-1996; U09122.
PR   07-JUN-1995; US-487032.
PR   01-APR-1996; US-630405.
PA   (ASTR ) ASTRA AB.
PI   Berglindh OT,  Smith D,  Mellgaerd BL;
DR   WPI; 97-052306/05.
DR   P-PSDB; W20805.
PT   Helicobacter pylori nucleic acid sequences and related
PT   polypeptide(s)  - useful for vaccines to treat or prevent H. pylori
PT   infection, and to detect Helicobacter
PS   Claim 9; Page 872-873; 1481pp; English.
CC   The present sequence encodes a H. pylori cytoplasmic protein involved in
CC   genome replication, transcription, recombination and repair.
CC   The protein may be used in a vaccine to prevent or treat H. pylori
CC   infection or to identify H. pylori polypeptide binding compounds,
CC   useful as potential H. pylori life cycle activators or inhibitors.
CC   The genomic sequence of H. pylori (ATCC 55679) was determined from
CC   overlapping contigs generated by mechanically shearing the bacterial
CC   DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC   and the predicted coding regions defined by computer evaluation. To
```

```
CC   identify likely H. pylori antigens for vaccine development, the amino
CC   acid sequences predicted from various ORF were analysed for significant
CC   homology to other known or exported membrane proteins. Having identified
CC   and determined the sequences of interest, particular regions can be
CC   isolated from H. pylori by PCR amplification for recombinant polypeptide
CC   production, e.g. in E. coli hosts.
SQ   Sequence   1215 BP;    354 A;    243 C;    270 G;    348 T;


alignment_scores:
            Quality:   74.00              Length:    266
              Ratio:  0.592                Gaps:     14
 Percent Similarity:  46.992   Percent Identity: 18.421

alignment_block:
 US-09-068-804-7 x T68058   ..

 Align seg 1/1  to: T68058  from: 1  to: 1215

     101 GlnAsnGluAsnLeuArgSerAlaPheSerAlaProThr...SerAlaLe 116
         |||   ::::::   ::::::   |||   |||::::::   :::||||| 
     388 CAAGTCCAGCACCACCATGCCCACTTTTTAGCGAGCGTCTTAGACGCATT 437

     116 uPheSerAlaSerProMetAlaGlnProArgThrThrIleSerAspAlaG 133
         |:::   ::::::   ::::::::::|||   :::   |||          
     438 ATTACAAGATCCGCATTTAAATCACCCCTTTATAGGTATT.......... 477

     133 luIleTrpAspMetValSerGlnAsnIleSerAlaIle......GlyAsp 147
           :::||||||       :::       :::   |||       ::::::  
     478 ..GTTTGGGACGGGAGTGGGGCTTATGAAAATAAGATTTATGGGGCGGAG 525

     148 SerTyrLeuGlyValTyrGluAsnVal......ValAlaValTyrThrAs 162
         ::::::::::|||       |||   :::      ||||||   :::   :: 
     526 TGTTTTGTGGGGGATTTGGAACGCATTGAAGAAGTGGCTAGATTTGAAGA 575

     162 pPheTyrGlnAlaPheSerAspIleLeuSerLysMetGlyGlyTrpLeuL 179
         :|||                                     |||||||  
     576 ATTT.....................................TGGCTTT 586

     179 euProGly............LysAspGlyAsnThrValLysLeuAspVal 191
         ||   |||            |||:::      :::   |||::::::  
     587 TAGGGGGGCAAAAAGCGATCAAAGAGCCTAGAAGATTGGTTTTAGAAATC 636

     192 ThrSerLeuLysAsnAspLeuAsnSerLeuValAsnLys........... 204
            :::||||||:::::::||||||   |||:::::::::          
     637 ...GCTTTAAAACACCAGCTCAACAAGCTTTTAAAACGCGTTCAAAAGCA 683

     205 ........................TyrAsnGlnIleAsnSerAsnThrV 213
                                 ::::::|||:::::::::::::::  
     684 TTTCAAAGAAGACGAATTAGGAATTTTTAAACAAATGCATGACAAAAAAA 733

     213 al..............................LeuPheProAlaGlnSer 219
         ::                           ||||||      :::      
     734 TTCAAAGCGTCGCCACCAATTCCATAGGGCGTTTGTTTGATATAGTGGCG 783

     220 .........GlySerGlyValLysValAlaThrGluAlaGluAlaArgGl 233
                 |||         :::::::::   |||||||||:::   ||  
     784 TTTAGTTTGGGCGTGGTGGGAACGATTAGTTTTGAAGCCGAGAGCGGGCA 833

     233 nTrpLeuSerGluLeuAsnLeuProAsnSerCysLeuLysSerTyrGlyS 250
         |   |||   :::|||:::|||::::::::::   :::      |||     
     834 GGTTTTAGAAAATCTAGCCCTACAAAGCGATGAGATCGCTTTTTAC.... 879

     250 erGlyTyrValValThrValAspLeuThrProLeuGlnLysMetValGln 266
               :::   ::::::   :::::::::   |||:::              
     880 ..CCTTTTGAAATCAAAAACAGCGTGGTGCGTTTGAAGGAATTT...... 921

     267 AspIleAspGlyLeuGlyAlaProGlyLysAspSerLysLeuGluMetAs 283
                                                             
     921 ..................................................... 921

     283 pAsnAlaLysTyrGlnAlaTrpGlnSerGlyPheLysAlaGlnGluGluA 300
                 |||||||||   :::                             

     922 .........TATCAAGCGTTTGAA..................... 936

     300 snMetLysThrThrLeuGlnThrLeuThrGlnLysTyrSerAsnAlaAsn 316
               :::   |||   :::|||   :::|||       |||:::      
     937 ........AAGGATTTGGGCGTTTTAGAACCCAAA...CGCATCGCTAAG 975

     317 SerLeuTyrAspAsnLeuValLysValLeuSerSerThrIleSerSer 332
           ::::::::::::||||||   ::::::::::::   |||::::::    
     976 AAATTTTTTAACAGCTTAGTAGAAATCATTACCGCTTTGATTGCGCCT 1023

seq_name: N_Geneseq_36:T68849

seq_documentation_block:
ID   T68849 standard; DNA; 4698 BP.
AC   T68849;
DT   30-JAN-1998  (first entry)
DE   Photorhabdus luminescens insect toxin TccB.
KW   Insecticide; insect; toxin; pest control; biological control;
KW   Photorhabdus luminescens; TccB; Southern corn rootworm;
KW   Colorado potato beetle; Western corn rootworm; meal worm;
KW   boll weevil; turf grub; Coleoptera; beet armyworm; black cutworm;
KW   cabbage looper; codling moth; corn earworm; European corn borer;
KW   tobacco hornworm; tobacco budworm; Lepidoptera; Hymenoptera;
KW   Diptera, Dictyoptera; Acarina; Homoptera; ds.
OS   Photorhabdus luminescens strain W-14 (ATCC 55397).
PN   WO9717432-A1.
PD   15-MAY-1997.
PF   06-NOV-1996; U18003.
PR   28-AUG-1996; US-705484.
PR   06-NOV-1995; US-007255.
PR   28-FEB-1996; US-608423.
PA   (WISC ) WISCONSIN ALUMNI RES FOUND.
PI   Blackburn MB,  Bowen DJ,  Ciche TA,  Ensign JC,  Fatig R;
PI   Ffrench-Constant RH,  Guo L,  Hey TD,  Merlo DJ,  Orr GL;
PI   Petell J,  Roberts JL,  Rocheleau TA, Schoonover S;
PI   Strickland JA;
DR   WPI; 97-281022/25.
DR   P-PSDB; W18305.
PT   Photorhabdus sp. insecticidal protein toxins and DNA encoding them -
PT   can be genetically engineered into insect larvae food and plants for
PT   insect control
PS   Claim 38; Page 239-245; 276pp; English.
CC   This genomic DNA sequence encodes the 185 kDa TccB insecticidal
CC   toxin protein (see W18305) of Photorhabdus luminescens W-14.
CC   Claimed toxins of P. luminescens (see W17871, W17884-89, W17899-900,
CC   W18301-06) can be produced by recombinant DNA methods and applied to,
CC   or genetically engineered into, insect larvae food and plants for
CC   insect control.  The Photorhabdus toxins are particularly effective
CC   against Southern corn rootworm, Colorado potato beetle, Western corn
CC   rootworm, meal worm, boll weevil and turf grub (Coleoptera), beet
CC   armyworm, black cutworm, cabbage looper, codling moth, corn earworm,
CC   European corn borer, tobacco hornworm and tobacco budworm
CC   (Lepidoptera), and are also active against insects of the orders
CC   Hymenoptera, Diptera, Dictyoptera, Acarina and Homoptera. (All
CC   claimed).
SQ   Sequence   4698 BP;   1217 A;    1023 C;    1202 G;    1256 T;


alignment_scores:
            Quality:   80.50              Length:    347
              Ratio:  0.440                Gaps:     16
 Percent Similarity:  52.738   Percent Identity: 19.597

alignment_block:
 US-09-068-804-7 x T68849   ..

 Align seg 1/1  to: T68849  from: 1  to: 4698

       2 LeuAsnIleGlnAsnTyrSerAlaSerProHisPro...GlyIleValAl 17
         |||::::::::::   |||:::|||:::   :::|||   ::::::::|| 
    3043 CTTTCGCTGCCGCTGTATGCTGCGCCTGTTGATCCGGTAGCGTTGTTGGC 3092

      17 aGluArgProGlnThrProSerAlaSerGluHisValGluThrAlaValV 34
```

```
    |:::|||:::|||:::   :::   ::::::   |||      |||:::: 
3093 TCAGCGTGCTCAGTCCGGCACGTTGACGAATGGCGTCAGTGGCGCCATGT 3142

  34 alProSerThrThrGluHisArgGlyThrAspIleIleSerLeuSerGln 50
     :::::      |||           |||::::::     |||:::::: 
3143 TGACGGTGGGTACG.........TTGACCAGTTTTGCTATGTTGCCGCGA 3183

  51 AlaAlaThrLysIle...................HisGlnAlaGlnGl 60
     |||   :::   :::                   ::::::
3184 GCTTACAGCGCCGTGGGTACGTTGACCAGTTTTGGTCAGAACCTGCTTAG 3233

  60 nThrLeuGlnSerThrProProIleSerGluGluAsnAsnAspGluArgT 77
         |||:::   :::         :::::::|||:::
3234 TTTGTTGGAACGTAGCGAACGAGCCTGTCAAGAAGAG............. 3270

  77 hrLeuAlaArgGlnGlnLeuThrSerSerLeuAsnAlaLeuAlaLysSer 93
     ||||||:::|||||||||                  |||:::::|||
3271 ..TTGGCGCAACAGCAACTG.................TTGGATATGTCC 3300

  94 GlyValSerLeuSerAlaGluGlnAsnGluAsnLeuArgSerAlaPheSe 110
     :::   :::::::::   :::|||::::::         ::::::::::
3301 AGCTATGCCATCACGTTGCAACAACAGGCG......CTGGATGGATTGGC 3344

 110 rAlaProThrSerAlaLeuPheSerAlaSerProMetAlaGlnProArgT 127
     :|||         ||||||:::::::::   :::   ||||||:::|||
3345 GGCAGATCGTCTGGCGCTGCTAGCTAGTCAGGCTACGGCACAACAGCGTC 3394

 127 hrThrIleSerAspAlaGluIleTrpAspMetValSerGlnAsnIleSer 143

3395 AT.............................................. 3396

 144 AlaIleGlyAspSerTyrLeuGlyValTyrGluAsnValValAlaValTy 160
             |||   |||      :::|||:::|||   ::::::
3397 .........GACCATTATTACACTCTGTATCAGAACAACATCTCCAGTGC 3437

 160 rThrAspPhe.........TyrGlnAlaPheSerAspIleLeuSerLysM 174
         ::::::           |||:::   ::::::   |||
3438 GGAACAACTGGTGATGGACACCCAAACGTCAGCACAATCCCTGATTTCTT 3487

 174 etGlyGlyTrpLeuLeuProGlyLysAspGlyAsnThrValLysLeuAsp 190
       :::          :::|||   :::   ::::::      |||·
3488 CTTCC.............ACTGGTGTACAAACTGCCAGTGGGGCACTGAAA 3525

 191 ValThrSerLeuLysAsnAspLeuAsnSerLeuValAsnLysTyrAsnGl 207
     |||::::::        :::|||:::::::      :::::|||:::
3526 GTGATCCCGAATATCTTTGGTTTGGCTGATGGCGGCTCGCGCTATGAAGG 3575

 207 nIleAsnSerAsnThrVal......LeuPheProAlaGlnSerGlySerG 222
      ::::::   :::::::::      |||   :::|||      :::::::
3576 AGTAACGGAAGCGATTGCCATCGGGTTAATGGCTGCCGGACAAGCCACCA 3625

 222 lyVal.........LysValAlaThrGluAlaGluAlaArg......... 232
     ::|||         ::::::||||||   ::::::   |||
3626 GCGTGGTGGCCGAGCGTCTGGCAACCACGGAGAATTACCGCCGCCGCCGT 3675

 233 ...GlnTrpLeuSerGluLeuAsnLeuProAsnSerCysLeuLysSerTy 248
        :::|||      :::   :::   :::::::|||
3676 GAAGAGTGGCAAATCCAATACCAGCAGGCACAGTCT.............. 3711

 248 rGlySerGlyTyrValValThrValAspLeuThrProLeuGlnLysMetV 265
                            ::::::   :::||||||||||
3712 ........................GAGGTCGACGCATTACAGAAA.... 3732

 265 alGlnAspIleAspGlyLeuGlyAlaProGlyLysAspSerLysLeuGlu 281
          :::::::|||:::|||::::::      |||::::::::::
3733 .....CAGTTGGATGCGCTGGCAGTGCGCGAGAAAGCAGCTCAAACTTCC 3777

 282 MetAspAsnAlaLysTyrGlnAlaTrpGlnSerGlyPheLysAlaGlnGl 298
     ::::::::::|||                              |||||||||::
3778 CTGCAACAGGCG.........................AAGGCACAGCA 3800

 298 uGluAsnMetLysThrThrLeuGlnThrLeuThrGlnLysTyrSerAsnA 315
     :   ::::::::::|||   |||      ||||||   ::::::::::::|
```

```
3801 GGTACAAATTCGGACCATGCTGACTTACTTAACTACTCGTTTCACCCAGG 3850

 315 laAsnSerLeuTyrAspAsnLeuValLysValLeuSerSer 328
     ||   :::|||||||:::   |||         |||||||:::
3851 CG...ACTCTGTACCAGTGGCTGAGTGGTCAATTATCCGCG 3888
```

seq_name: N_Geneseq_36:X12983

seq_documentation_block:
ID X12983 standard; DNA; 3169 BP.
AC X12983;
DT 19-MAR-1999 (first entry)
DE Enterococcus faecalis genome contig SEQ ID NO:46.
KW Enterococcus faecalis; contig; detection; Enterococcal infection;
KW vaccine; attenuation; computer readable medium; ds.
OS Enterococcus faecalis.
PN WO9850555-A2.
PD 12-NOV-1998.
PF 04-MAY-1998; U08985.
PR 14-NOV-1997; US-066009.
PR 06-MAY-1997; US-044031.
PR 16-MAY-1997; US-046655.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Dillon PJ, Kunsch CA;
DR WPI; 99-045171/04.
PT New isolated Enterococcus faecalis polynucleotides and polypeptides
PT - used to develop products for the detection of Enterococcus and for
PT use in vaccines for prevention or attenuation of Enterococcus
PT infection.
PS Claim 1; Page 427-429; 2084pp; English.
CC A computer readable medium has been developed which has recorded on it
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC X12938 to X13919 represent these nucleotide sequences which are primary
CC nucleotide sequences, also known as contigs. The computer-based system
CC can identify fragments of the Enterococcus faecalis genome with
CC commercial importance. The products can be used to detect the presence
CC of Enterococcus faecalis in samples. They can also be used for
CC diagnosing Enterococcal infection in an animal and monitoring
CC progression of disease, and for identifying agents which can be used to
CC modulate the growth or pathogenicity of Enterococcus faecalis, or
CC another related organism, in vivo or in vitro. In particular the
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC can be used in vaccines to prevent or attenuate an Enterococcal
CC infection.
SQ Sequence 3169 BP; 1100 A; 466 C; 611 G; 977 T;

alignment_scores:
Quality: 78.00 Length: 227
Ratio: 0.672 Gaps: 12
Percent Similarity: 51.101 Percent Identity: 19.824

alignment_block:
US-09-068-804-7 x X12983 ..

Align seg 1/1 to: X12983 from: 1 to: 3169

```
    109 PheSerAlaProThrSerAlaLeuPheSerAlaSerProMet..... 122
        |||:::      |||   ::::::|||   :::|||::::::
    132 TTTGGGTTTCGAACACAGGATATTTTTTATAGTTCCACTATTGAAACAGG 181

    123 .AlaGlnProArgThrThrIleSerAspAlaGluIleTrpAspMetValS 139
         :::|||||          |||   |||:::      ::::::::::::
    182 TGTACAGCCAGGAGAATTAATTTTTGACACGATTAAAAGGGAGCTAACCA 231

    139 erGlnAsnIleSerAlaIleGlyAspSerTyrLeuGlyValTyrGluAsn 155
        ::|||
    232 ATCAACCT..................................... 239

    156 ValValAlaValTyr.........ThrAspPheTyrGlnAlaPheSerAs 169
           |||   :::|||           ::::::||||||:::::::::
    240 ...GTCATGCTGTATTTTTTATCTGATCATTATTATCAGAGTATTCCA.. 284
```

```
     169 pIleLeuSerLysMetGly...GlyTrpLeuLeuProGlyLysAspGlyA 185
         |||:::   ||||||   :::|||:::|||:::::|||:::
     285 .TGTTTAAACGAAATGGGTGCTTCTTGGATGCTATCAGATAAACAC...T 330

     185 snThrValLysLeuAspValThrSerLeuLysAsnAspLeuAsnSerLeu 201
           ::::::   |||:::      |||:::|||   |||:::::::::::::
     331 ATCCAATTGCGTTAAATAATTTTTCTATGAAA...GATATGAAAGGTGTT 377

     202 ValAsnLysTyrAsnGlnIleAsnSerAsnThrValLeuPheProAlaGl 218
         :::
     378 ATT..................................................... 380

     218 nSerGlySerGlyValLysValAlaThrGluAlaGluAlaArgGlnTrpL 235
               |||:::   :::::::|||::::::   :::::::::
     381 .......AGTAGTGAGCGTTTGGCGATTGCGTTTAATGATAAGACAAGTA 423

     235 euSerGluLeuAsnLeuProAsnSerCysLeuLysSerTyrGlySerGly 251
           :::|||:::|||             ||||||
     424 CTAATGAAATAAAC............TGCTTA..................... 443

     252 TyrValValThrValAspLeuThrProLeuGlnLysMetValGlnAspIl 268
                                 |||:::|||:::   :::|||::
     444 ..........................TTAAAAAAGTTATCTCATGACAC 466

     268 eAspGlyLeuGlyAlaProGlyLysAspSerLysLeuGluMetAspAsnA 285
         :|||      :::::::|||:::   :::   ::::::|||:::::::
     467 AGATGTACAAGCTGAGCCAGATTTTGAACTAAACGTTGAGAAAAATATTC 516

     285 laLysTyrGlnAla..................TrpGlnSerGlyPheLys 295
           :::|||:::                    :::|||::::::::::
     517 AGCCATTCCAAAATAAACTGACACAACTGATTCGGCAAGCAAGTTATTTA 566

     296 AlaGlnGluGluAsn......MetLysThrThrLeuGlnThrLeuThrGl 310
           ::::::|||:::             |||   |||   |||      ::
     567 AAACCTGATGAAAAGGGATATTTTGAAACGRTTCTTAGCACTCATCGTCC 616

     310 nLysTyrSerAsnAlaAsnSerLeuTyrAsp 320
         :   |||::::::|||:::::::::||||||
     617 TGTGTATGGCACAGCAAAAGGGGTCTATGAC 647
```

```
seq_name: N_Geneseq_36:V71025

seq_documentation_block:
ID   V71025 standard; cDNA; 1824 BP.
AC   V71025;
DT   08-FEB-1999  (first entry)
DE   p38-green flourescent protein fusion product.
KW   Human; p38 gene; fusion protein; green fluorescent protein; GFP;
KW   intracellular signalling; chimera; ss.
OS   Chimeric - Aequorea victoria.
OS   Chimeric - Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             1. .1824
FT                   /*tag=  a
PN   WO9845704-A2.
PD   15-OCT-1998.
PF   07-APR-1998; DK0145.
PR   07-APR-1997; DK-000392.
PA   (NOVO ) NOVO-NORDISK AS.
PI   Kasper A, Petersen Bjorn S, Scudder K, Thastrup O,
PI   Tullin S;
DR   WPI; 98-594491/50.
DR   P-PSDB; W85010.
PT   Determining effect on signalling pathways in live cells from
PT   redistribution of luminophores - specifically fusions of green
PT   fluorescent protein with a signalling component, and new apparatus,
PT   particularly for identifying toxins and potential therapeutic agents
PS   Claim 63; Pages 87-89; 326pp; English.
CC   The present sequence encodes a human p38-green flourescent
CC   protein fusion product. The fusion protein is used in an assay
CC   that exemplifies the invention. The specification describes how
CC   quantitative information about the influence of a molecule on a cellular
CC   response is obtained by recording the variation, caused by the molecule,
```

```
CC   on mechanically intact living cells, in the spatially distributed light
CC   emitted from a luminophore present in the cells. The variation in light
CC   emission is processed to provide information that correlates spatial
CC   distribution to the degree of the molecule. The method is used to
CC   identify agents that (in)directly affect intracellular signalling,
CC   especially to screen for potential therapeutic agents or toxins, and
CC   to identify new drug targets.
SQ   Sequence   1824 BP;    476 A;    482 C;    478 G;    388 T;


alignment_scores:
           Quality:   75.00              Length:    197
             Ratio:  0.743                Gaps:     11
 Percent Similarity:  51.269   Percent Identity: 22.335

alignment_block:
 US-09-068-804-7 x V71025   ..

 Align seg 1/1  to: V71025  from: 1  to: 1824

      11 ProHisProGlyIleValAlaGluArgProGlnThrProSerAlaSerGl 27
         |||:::   |||   |||      |||   :::   |||::::::::::::
     327 CCGCGCCGAGGTGAAGTT......CGAGGGCGACACCCTGGTGAACCGCA 370

      27 uHisValGluThrAlaValValProSerThrThrGluHisArgGlyThrA 44
         :::::::|||         :::::::::      :::|||   |||::::
     371 TCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCAC 420

      44 spIleIleSerLeuSerGlnAlaAlaThrLysIleHisGlnAlaGlnGln 60
         ::               |||   :::   ::::::::::::::::::|||
     421 AAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAA 470

      61 Thr.....................LeuGlnSerThrPro........... 66
         :::                    ||||||::::::|||
     471 GCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGG 520

      67 ......................ProIleSerGluGluAsnAsnAsp.... 74
                               |||::::::::::|||:::   :::
     521 ACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGC 570

      75 .....GluArgThrLeuAlaArgGlnGlnLeuThr.SerSerLeuAsnAl 89
               |||:::   |||||||||:::|||::: |||:::   :::||
     571 GACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGC 620

      89 aLeuAlaLys..................................... 92
         ||||:::|||
     621 CCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGT 670

      93 ..........SerGlyValSerLeuSerAlaGluGln......AsnGlu 103
                  :::|||::::::::|||:::   ::::::      :::
     671 TCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTCC 720

     104 AsnLeuArgSerAlaPheSerAlaProThrSerAlaLeuPheSerAlaSe 120
         :::||||||                                  |||   ::
     721 GGACTCAGA..............................TCTCGAGG 737

     120 rProMetAlaGlnProArgThrThrIleSerAspAlaGlu.......... 133
         :   |||:::|||   |||:::|||:::         |||
     738 GAAAATGTCTCAGGAGAGGCCCACGTTCTACCGGCAGGAGCTGAACAAGA 787

     134 ..IleTrpAspMetValSer......GlnAsnIleSerAlaIleGlyAsp 147
           ||||||:::::            ||||||:::|||:::::|||:::
     788 CAATCTGGGAGGTGCCCGAGCGTTACCAGAACCTGTCTCCAGTGGGCTCT 837

     148 SerTyrLeuGlyValTyrGluAsnValValAlaValTyr 160
               |||:::|||   :::|||   |||::::::
     838 .........GGCGCCTATGGCTCTGTGTGTGCTGCTTTT 867
```

```
seq_name: N_Geneseq_36:X09010

seq_documentation_block:
ID   X09010 standard; cDNA; 1272 BP.
AC   X09010;
```

```
DT   14-JUN-1999  (first entry)
DE   Brn-3a polynucleotide.
KW   Brn-3a; Bcl-2; neurons; neuronal cells; apoptosis; cell death; CNS;
KW   PNS; central nervous system; parasympathetic nervous system;
KW   development; injury; neurotrophic factor; nerve growth factor; NGF;
KW   ciliary neurotrophic factor; CNTF; brain-derived neurotrophic factor;
KW   BNTF; neurotrophin; NT-3; NT-4; NT-5; neurodegenerative disease;
KW   familial dysautonomia; infantile muscular dystrophy;
KW   Parkinson's disease; Alzheimer's disease; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             1. .1272
FT                   /*tag= a
FT                   /product= Brn_3a_polypeptide
PN   WO9905272-A1.
PD   04-FEB-1999.
PF   27-JUL-1998; G02228.
PR   10-DEC-1997; US-988476.
PR   25-JUL-1997; GB-015823.
PA   (UNLO ) UNIV COLLEGE LONDON.
PI   Latchman DS, Smith MD;
DR   WPI; 99-142928/12.
DR   P-PSDB; W96262.
PT   New polypeptide comprising transcription factor Brn-3a, or its
PT   derivative - useful for treating nervous system diseases, preventing
PT   cellular apoptosis and increasing nerve regeneration following
PT   neuronal damage
PS   Disclosure; Page 61-62; 68pp; English.
CC   Over expression of transcription factor Brn-3a can protect neuronal
CC   cells from apoptosis. Brn-3a also specifically activates expression
CC   of the BCl-2 gene in neuronal cells and this activation is mediated
CC   via a Brn-3a response element in the 5' regulatory region of the
CC   Bcl-2 gene. Both the anti-apoptopic effect of Brn-3a and its ability
CC   to activate expression of Bcl-2 are mediated by the N-terminal
CC   domain of Brn-3a. Members of the Bcl-2 family perform critical roles
CC   in the regulation of selective apoptosis during development of the
CC   nervous system.  The stimulation of Bcl-2 expression by Brn-3a in a
CC   neuron specific manner and consequent protection of neuronal cells
CC   from apoptosis suggests that Brn-3a may co-ordinate some aspects of
CC   neuronal reorganisation during development or following injury.  The
CC   elevation of Brn-3a expression by either pharmacological means
CC   (compositions comprising one other therapeutic polypeptide e.g.
CC   neurotrophic factors, nerve growth factor (NGF), ciliary neurotrophic
CC   factor (CNTF), brain-derived neurotrophic factor (BNTF), and
CC   neurotrophins NT-3 and NT-4/5) or gene therapy may represent a
CC   method for treating human diseases associated with excessive
CC   neuronal cell death and/or lack of nerve regeneration, especially
CC   neurodegenerative diseases such as familial dysautonomia and
CC   infantile muscular dystrophy, and Parkinson's and Alzheimer's
CC   disease.
SQ   Sequence   1272 BP;    202 A;    468 C;    452 G;    150 T;


alignment_scores:
           Quality:   73.00            Length:    202
             Ratio:  0.716              Gaps:      6
 Percent Similarity:  50.495   Percent Identity: 17.822

alignment_block:
 US-09-068-804-7 x X09010   ..

 Align seg 1/1  to: X09010  from: 1  to: 1272

      11 ProHisProGlyIleValAlaGluArgProGlnThrProSerAlaSerGl 27
         ||||||||||||:::||||||:::   ::::::      :::|||:::::
     655 CCGCACCCCGGGCTGGTGGCGGCGGCGGCGCACCACGGCGCGGCAGCGGC 704

      27 uHisValGluThrAlaValValProSerThrThrGluHisArgGlyThrA 44
         :   ::::::::::|||:::   :::   :::::::::      :::::::
     705 AGCGGCGGCGGCGGCGGCCGGGCAGGTGGCAGCGGCATCGGCGGCGGCGG 754

      44 spIleIleSerLeuSerGlnAlaAlaThrLysIleHisGlnAlaGlnGln 60
         ::::::::::   :::

     755 CCGTGGTGGGCGCAGCG..............................  771

      61 ThrLeuGlnSerThrProProIleSerGluGluAsnAsnAspGluArgTh 77
                 :::   ::::::|||::::::   ::::::|||   |||
     772 .........GGCCTGGCGTCCATCTGCGACTCGGACACGGACCCGCGCGA 812

      77 rLeu.......................................... 78
          |||
     813 GCTCGAGGCGTTCGCGGAGCGCTTCAAGCAGCGGCGCATCAAGCTGGGCG 862

      79 ..AlaArgGlnGlnLeuThrSerSerLeuAsnAlaLeuAlaLysSerGly 94
           ::::::   ::::::   |||:::|||::::::|||      :::|||
     863 TGACGCAGGCCGACGTGGGCTCGGCGCTGGCCAACCTCAAGATCCCGGGC 912

      95 Val.....................................SerLe 97
         |||                                     :::||
     913 GTGGGCTCACTCAGCCAGAGCACCATCTGCAGGTTCGAGTCGCTCACGCT 962

      97 uSerAlaGluGlnAsnGluAsnLeuArgSerAlaPheSerAla..... 111
         ||||   ::::::      :::|||::::::   :::   |||
     963 CTCGCACAACAACATGATCGCGCTCAAGCCCATCCTGCAGGCGTGGCTCG 1012

     112 ..............................ProThrSerAlaLeuPhe 117
                                        :::::::::||||||
    1013 AGGAGGCCGAGGGCGCCCAGCGCGAGAAAATGAACAAGCCTGAGCTCTTC 1062

     118 SerAlaSerProMetAlaGlnProArgThrThrIleSerAspAlaGluIl 134
         :::::::::   :::   :::   ||||||:::|||:::::::::|||
    1063 AACGGCGGCGAGAAGAAGCGCAAGCGGACTTCCATCGCCGCGCCCGAGAA 1112

     134 eTrpAspMet.............................ValSerGlnA 141
          :::::::::                             |||::::
    1113 GCGCTCCCTCGAGGCCTACTTCGCCGTGCAGCCCCGGCCCTCGTCCGAGA 1162

     141 snIleSerAlaIleGlyAspSerTyrLeuGlyValTyrGluAsnValVal 157
         ::|||:::||||||:::::      |||:::::      |||||||||
    1163 AGATCGCCGCCATCGCCGAGAAA...CTGGACCTCAAAAAGAACGTGGTG 1209

     158 AlaVal 159
            |||
    1210 CGGGTG 1215

seq_name: N_Geneseq_36:V71180

seq_documentation_block:
ID   V71180 standard; cDNA; 2143 BP.
AC   V71180;
DT   12-FEB-1999  (first entry)
DE   Clone 1711346IH, the PS108 gene contig full length sequence.
KW   PS108 gene; prostate disease; benign prostatic hyperplasia; BPH;
KW   prostatitis; prostatic intraepithelial neoplasia; PIN; cancer;
KW   drug screening; gene therapy; ss.
OS   Homo sapiens.
PN   WO9850567-A1.
PD   12-NOV-1998.
PF   01-MAY-1998; U08930.
PR   02-MAY-1997; US-850713.
PA   (ABBO ) ABBOTT LAB.
PI   Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN,
PI   Gordon J, Granados EN, Hodges SC, Klass MR, Kratochvil JD,
PI   Roberts-Rapp L, Russell JC, Stroupe SD;
DR   WPI; 99-034731/03.
PT   New isolated prostate-specific polynucleotides - used to develop
PT   products for the diagnosis and treatment of prostate diseases, e.g.
PT   benign hyperplasia, prostatic or prostate cancer
PS   Claim 1; Fig 1A-E; 122pp; English.
CC   The present sequence represents the full length contig of the PS108
CC   gene, as represented by clone 1711346IH. This clone is the contig of
CC   overlapping clones V71166-79. The clone sequences are PS108
CC   gene-specific. They are used in the method of the invention. The
CC   specification describes a method for detecting the presence of a
CC   target PS108 polynucleotide in a test sample. The method comprises
CC   contacting the test sample with at least 1 PS108-specific polynucleotide
```

```
CC   or complement, and detecting the presence of the target PS108
CC   polynucleotide. The products can be used for detecting, diagnosing,
CC   staging, monitoring, prognosticating, in vivo imaging, preventing or
CC   treating, or determining predisposition to diseases or conditions of
CC   the prostate such as benign prostatic hyperplasia (BPH), prostatitis,
CC   prostatic intraepithelial neoplasia (PIN) and cancer. In particular the
CC   products can be used in drug screening and gene therapy.
SQ   Sequence   2143 BP;    418 A;    621 C;    563 G;    541 T;


alignment_scores:
           Quality:   75.50            Length:    100
             Ratio:  1.079              Gaps:      4
 Percent Similarity:  70.000   Percent Identity: 26.000

alignment_block:
 US-09-068-804-7 x V71180   ..

 Align seg 1/1  to: V71180  from: 1  to: 2143

      77 ThrLeuAlaArgGlnGlnLeuThrSerSerLeuAsnAlaLeuAlaLysSe 93
         ||||||:::|||         |||:::   ::::::||||||:::|||
    1767 ACTCTCTCTAGG.........ACTGGGCTGATGAAGGCACTGCCCAAA.. 1805

      93 rGlyValSerLeuSerAlaGluGlnAsnGluAsnLeuArgSerAlaPheS 110
                 :::|||:::   :::|||      |||   ::::::|||:
    1806 .........ATTTCCCCTACCCCCAACTTTCCCCTACCCCCAACTTTCC 1845

     110 erAlaProThrSerAlaLeuPheSerAlaSerProMetAlaGlnProArg 126
         ::::::::::::::::::||||||:::|||::::::   ::::::::::::
    1846 CCACCAGCTCCACAACCCTGTTTGGAGCTACTGCAGGACCAGAAGCACAA 1895

     127 ThrThrIleSerAspAlaGluIleTrpAspMetValSerGlnAsnIleSe 143
         :::::::::|||:::|||   :::   ::::::   :::|||:::|||||
    1896 AGTGCGGTTTCCCAAGCCTTTGTC...CATCTCAGCCCCCAGAGTATATC 1942

     143 rAlaIleGlyAspSerTyrLeuGlyValTyrGluAsnValValAlaValT 160
         |:::::::|||:::|||:::      :::   |||:::   ::::::
    1943 TGTGCTTGGGGAATCTCACACAGAAACTCAGGAGCACCCCCTGCCT.... 1988

     160 yrThrAspPheTyrGlnAlaPheSerAspIleLeuSerLysMetGlyGly 176
                       ::::::   ::::::||||||      ||||||
    1989 .................GAGCTAAGGGAGGTCTTATCTCTCAGGGGGGGT 2021

seq_name: N_Geneseq_36:V71181

seq_documentation_block:
ID   V71181 standard; cDNA; 2152 BP.
AC   V71181;
DT   12-FEB-1999  (first entry)
DE   Consensus sequence of the PS108 gene derived from overlapping clones.
KW   PS108 gene; prostate disease; benign prostatic hyperplasia; BPH;
KW   prostatitis; prostatic intraepithelial neoplasia; PIN; cancer;
KW   drug screening; gene therapy; ss.
OS   Synthetic.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             2. .769
FT                   /*tag=  a
PN   WO9850567-A1.
PD   12-NOV-1998.
PF   01-MAY-1998; U08930.
PR   02-MAY-1997; US-850713.
PA   (ABBO ) ABBOTT LAB.
PI   Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN,
PI   Gordon J, Granados EN, Hodges SC, Klass MR, Kratochvil JD,
PI   Roberts-Rapp L, Russell JC, Stroupe SD;
DR   WPI; 99-034731/03.
DR   P-PSDB; W85068.
PT   New isolated prostate-specific polynucleotides - used to develop
PT   products for the diagnosis and treatment of prostate diseases, e.g.
PT   benign hyperplasia, prostatic or prostate cancer
PS   Claim 1; Fig 1A-E; 122pp; English.
```

```
CC   The present sequence represents the consensus sequence for a PS108
CC   gene. The sequence is derived from overlapping clones V71166-79. The
CC   clone sequences are PS108 gene-specific. They are used in the method
CC   of the invention. The specification describes a method for detecting the
CC   presence of a target PS108 polynucleotide in a test sample. The method
CC   comprises contacting the test sample with at least 1 PS108-specific
CC   polynucleotide or complement, and detecting the presence of the target
CC   PS108 polynucleotide. The products can be used for detecting, diagnosing,
CC   staging, monitoring, prognosticating, in vivo imaging, preventing or
CC   treating, or determining predisposition to diseases or conditions of
CC   the prostate such as benign prostatic hyperplasia (BPH), prostatitis,
CC   prostatic intraepithelial neoplasia (PIN) and cancer. In particular the
CC   products can be used in drug screening and gene therapy.
SQ   Sequence   2152 BP;    419 A;    622 C;    569 G;    542 T;


alignment_scores:
           Quality:   75.50            Length:    100
             Ratio:  1.079              Gaps:      4
 Percent Similarity:  70.000   Percent Identity: 26.000

alignment_block:
 US-09-068-804-7 x V71181   ..

 Align seg 1/1  to: V71181  from: 1  to: 2152

      77 ThrLeuAlaArgGlnGlnLeuThrSerSerLeuAsnAlaLeuAlaLysSe 93
         ||||||:::|||         |||:::   ::::::||||||:::|||
    1776 ACTCTCTCTAGG.........ACTGGGCTGATGAAGGCACTGCCCAAA.. 1814

      93 rGlyValSerLeuSerAlaGluGlnAsnGluAsnLeuArgSerAlaPheS 110
                 :::|||:::   :::|||      |||   ::::::|||:
    1815 .........ATTTCCCCTACCCCCAACTTTCCCCTACCCCCAACTTTCC 1854

     110 erAlaProThrSerAlaLeuPheSerAlaSerProMetAlaGlnProArg 126
         ::::::::::::::::::||||||:::|||::::::   ::::::::::::
    1855 CCACCAGCTCCACAACCCTGTTTGGAGCTACTGCAGGACCAGAAGCACAA 1904

     127 ThrThrIleSerAspAlaGluIleTrpAspMetValSerGlnAsnIleSe 143
         :::::::::|||:::|||   :::   ::::::   :::|||:::|||||
    1905 AGTGCGGTTTCCCAAGCCTTTGTC...CATCTCAGCCCCCAGAGTATATC 1951

     143 rAlaIleGlyAspSerTyrLeuGlyValTyrGluAsnValValAlaValT 160
         |:::::::|||:::|||:::      :::   |||:::   ::::::
    1952 TGTGCTTGGGGAATCTCACACAGAAACTCAGGAGCACCCCCTGCCT.... 1997

     160 yrThrAspPheTyrGlnAlaPheSerAspIleLeuSerLysMetGlyGly 176
                       ::::::   ::::::||||||      ||||||
    1998 .................GAGCTAAGGGAGGTCTTATCTCTCAGGGGGGGT 2030

seq_name: N_Geneseq_36:V29926

seq_documentation_block:
ID   V29926 standard; DNA; 4832 BP.
AC   V29926;
DT   07-AUG-1998  (first entry)
DE   tccB gene from the tcc genomic region encoding protein TccB.
KW   TcaA; Photorhabdus luminescens W-14; nematode; symbiotic;
KW   Heterorhabditis; tca; tcb; tcc; tcd; insecticidal activity; toxin;
KW   Lepidoptera; Coleoptera; Hymenoptera; Diptera; Dictyoptera; Acarina;
KW   Homoptera; Southern; Western corn rootworm; Colorado potato beetle;
KW   mealworm; boll weevil; turf grub; beetle armyworm; black cutworm;
KW   cabbage looper; codling moth; corn earworm; European corn borer;
KW   Tobacco hornworm; budworm; ds.
OS   Photorhabdus luminescens.
FH   Key             Location/Qualifiers
FT   CDS             1. .4698
FT                   /*tag=  a
FT                   /product= TccB
PN   WO9808932-A1.
PD   05-MAR-1998.
PF   05-MAY-1997; U07657.
PR   06-NOV-1996; WO-U18003.
```

```
PR   28-AUG-1996; US-705484.
PR   06-NOV-1996; US-743699.
PA   (DOWC ) DOWELANCO.
PA   (WISC ) WISCONSIN ALUMNI RES FOUND.
PI   Blackburn MB, Bowen DJ, Ciche TA, Ensign JC, Fatig R,
PI   Ffrench-Constant RH, Guo L, Hey TD, Merlo DJ, Orr GL,
PI   Petell J, Roberts JL, Rocheleau TA, Schoonover S,
PI   Strickland JA, Sukhapinda K;
DR   WPI; 98-179427/16.
DR   P-PSDB; W56568.
PT   Isolated toxins from Photorhabdus luminescens strains - useful for
PT   control of insect pests
PS   Claim 38; Pages 269-276; 321pp; English.
CC   The present sequence encodes a protein named TccB of the bacterium
CC   Photorhabdus luminescens (W-14). This is a symbiotic bacterium of the
CC   nematodes of the Heterorhabditis genus. The bacterium has at least 4
CC   distinct genomic regions, tca, tcb, tcc, and tcd. Peptide products are
CC   produced from these regions that are associated with insecticidal
CC   activity. The native toxins are secreted proteins. The proteins are
CC   toxic to insects upon exposure and especially when ingested. The
CC   nucleic acid sequence can be used to produce transgenic plants,
CC   baculoviruses or microbial hosts for toxin production.They can be used
CC   to control insects pests from the Lepidoptera, Coleoptera, Hymenoptera,
CC   Diptera, Dictyoptera, Acarina or Homoptera orders, especially the
CC   Southern or Western corn rootworm, Colorado potato beetle, mealworm,
CC   boll weevil, turf grub, beetle armyworm, black cutworm, cabbage looper,
CC   codling moth, corn earworm, European corn borer or tobacco hornworm
CC   or budworm.
SQ   Sequence   4832 BP;    1259 A;    1061 C;    1224 G;    1288 T;


alignment_scores:
            Quality:   79.50              Length:    357
              Ratio:  0.447               Gaps:      15
 Percent Similarity:  49.860   Percent Identity: 18.768

alignment_block:
 US-09-068-804-7 x V29926   ..

 Align seg 1/1  to: V29926  from: 1  to: 4832

       2 LeuAsnIleGlnAsnTyrSerAlaSerProHisPro...GlyIleValAl 17
         |||:::::::::   |||:::|||:::   :::|||   :::::::::|| 
    3043 CTTTCGCTGCCGCTGTATGCTGCGCCTGTTGATCCGGTAGCGTTGTTGGC 3092

      17 aGluArgProGlnThrProSerAlaSerGluHisValGluThrAlaValV 34
         |:::|||:::|||:::   :::   ::::::   |||       |||:::: 
    3093 TCAGCGTGCTCAGTCCGGCACGTTGACGAATGGCGTCAGTGGCGCCATGT 3142

      34 alPro..................................................... 35
         :::::
    3143 TGACGGTGCCGCCATACCGTTTCAGCGCTATGTTGCCGCGAGCTTACAGC 3192

      36 ......SerThrThrGluHisArgGlyThrAspIleIleSerLeuSerGl 50
               :::|||            |||   :::::::::||||||   :: 
    3193 GCCGTGGGTACGTTGACCAGTTTTGGTCAGAACCTGCTTAGTTTGTTGGA 3242

      50 nAlaAlaThrLysIleHisGlnAlaGlnGlnThrLeuGlnSerThrProP 67
         :   :::   :::      |||::::::
    3243 ACGTAGCGAACGAGCCTGTCAAGAAGAG..... 3270

      67 roIleSerGluGluAsnAsnAspGluArgThrLeuAlaArgGlnGlnLeu 83
                                       ||||||:::|||||||||| 
    3271 ...............................TTGGCGCAACAGCAACTG 3288

      84 ThrSerSerLeuAsnAlaLeuAlaLysSerGlyValSerLeuSerAlaGl 100
                        |||:::::::|||:::   :::::::::   :: 
    3289 ..................TTGGATATGTCCAGCTATGCCATCACGTTGCA 3320

     100 uGlnAsnGluAsnLeuArgSerAlaPheSerAlaProThrSerAlaLeuP 117
         :|||::::::         ::::::::::::|||         ||||||: 
    3321 ACAACAGGCG......CTGGATGGATTGGCGGCAGATCGTCTGGCGCTGC 3364

     117 heSerAlaSerProMetAlaGlnProArgThrThrIleSerAspAlaGlu 133
         ::::::::   :::   ||||||:::|||
    3365 TAGCTAGTCAGGCTACGGCACAACAGCGTCAT..... 3396

     134 IleTrpAspMetValSerGlnAsnIleSerAlaIleGlyAspSerTyrLe 150
                                                |||   |||
    3397 .....................................GACCATTATTA 3407

     150 uGlyValTyrGluAsnValValAlaValTyrThrAspPhe.........I 164
             :::|||:::|||   ::::::         ::::::
    3408 CACTCTGTATCAGAACAACATCTCCAGTGCGGAACAACTGGTGATGGACA 3457

     164 yrGlnAlaPheSerAspIleLeuSerLysMetGlyGlyTrpLeuLeuPro 180
           |||:::   ::::::   |||         :::           :::
    3458 CCCAAACGTCAGCACAATCCCTGATTTCTTCTTCC............ACT 3495

     181 GlyLysAspGlyAsnThrValLysLeuAspValThrSerLeuLysAsnAs 197
         |||   :::   ::::::      |||   |||::::::          :: 
    3496 GGTGTACAAACTGCCAGTGGGGCACTGAAAGTGATCCCGAATATCTTTGG 3545

     197 pLeuAsnSerLeuValAsnLysTyrAsnGlnIleAsnSerAsnThrVal. 213
         :|||::::::      :::::::|||:::   ::::::   :::::::::
    3546 TTTGGCTGATGGCGGCTCGCGCTATGAAGGAGTAACGGAAGCGATTGCCA 3595

     214 .....LeuProProAlaGlnSerGlySerGlyVal.........LysVal 225
               |||   :::|||       ::::::::::|||           ::::::
    3596 TCGGGTTAATGGCTGCCGGACAAGCCACCAGCGTGGTGGCCGAGCGTCTG 3645

     226 AlaThrGluAlaGluAlaArg............GlnTrpLeuSerGluLe 238
         ||||||   ::::::   |||            :::|||      :::
    3646 GCAACCACGGAGAATTACCGCCGCCGCCGTGAAGAGTGGCAAATCCAATA 3695

     238 uAsnLeuProAsnSerCysLeuLysSerTyrGlySerGlyTyrValValT 255
          :::   ::::::|||
    3696 CCAGCAGGCACAGTCT................................ 3711

     255 hrValAspLeuThrProLeuGlnLysMetValGlnAspIleAspGlyLeu 271
               ::::::   :::|||||||||         ::::::|||:::|||
    3712 .....GAGGTCGACGCATTACAGAAA.........CAGTTGGATGCGCTG 3747

     272 GlyAlaProGlyLysAspSerLysLeuGluMetAspAsnAlaLysTyrGl 288
         ::::::       |||:::::::::       ::::::::|||
    3748 GCAGTGCGCGAGAAAGCAGCTCAAACTTCCCTGCAACAGGCG........ 3789

     288 nAlaTrpGlnSerGlyPheLysAlaGlnGluGluAsnMetLysThrThrL 305
                         |||||||||:::   :::::::::|||   |
    3790 ..................AAGGCACAGCAGGTACAAATTCGGACCATGC 3820

     305 euGlnThrLeuThrGlnLysTyrSerAsnAlaAsnSerLeuTyrAspAsn 321
         ||       ||||||   :::::::::::::|||   :::|||||||:::
    3821 TGACTTACTTAACTACTCGTTTCACCCAGGCG...ACTCTGTACCAGTGG 3867

     322 LeuValLysValLeuSerSer 328
         |||          ||||||:::
    3868 CTGAGTGGTCAATTATCCGCG 3888
```